UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Departamento de Física de Partículas
Instituto Galego de Física de Altas Enerxías

TESE DE DOUTORAMENTO

Search for the rare decays $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$ with 1 fb^{-1} at LHCb

Presentada por:
Xabier Cid Vidal

Dirixida por:
Jose Ángel Hernando Morata
Bernardo Adeva Andany

Data:
19 de Xullo de 2012

Tribunal:
Dr. Pierluigi Campana (Presidente)
Dr. Cibrán Santamarina (Secretario)
Prof. Carlos Pajares
Dr. Thomas Ruf
Dr. Frederic Teubert

D. BERNARDO ADEVA ANDANY, Catedrático de Física Atómica, Molecular e Nuclear da Universidade de Santiago de Compostela e D. JOSE ÁNGEL HERNANDO MORATA, Profesor Titular de Física Atómica, Molecular e Nuclear da Universidade de Santiago de Compostela, informan:

Que a memoria titulada "Search for the rare decays $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$ with 1 fb^{-1} at LHCb" foi realizada por Xabier Cid Vidal baixo a súa dirección e que constitúe a Tese que presenta para optar ao Grao de Doutor en Física.

Santiago de Compostela, a 19 de Xullo de 2012.

Asdo. Bernardo Adeva Andany Asdo. Jose Ángel Hernando Morata Asdo. Xabier Cid Vidal

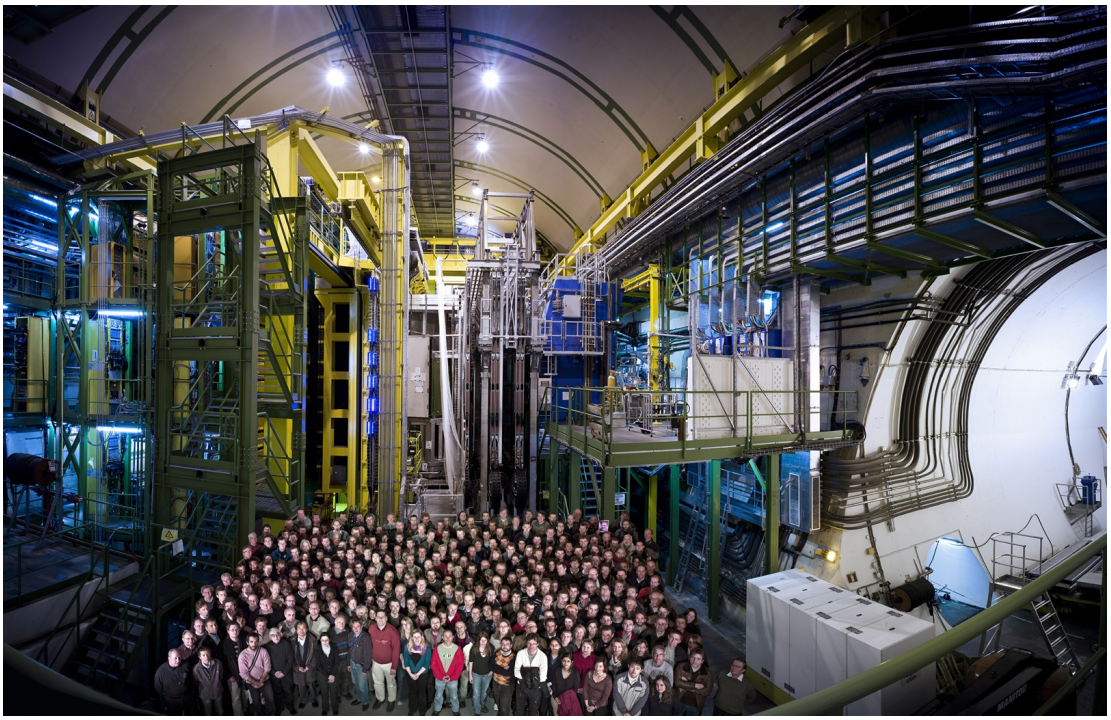

Figure 1: The LHCb collaboration

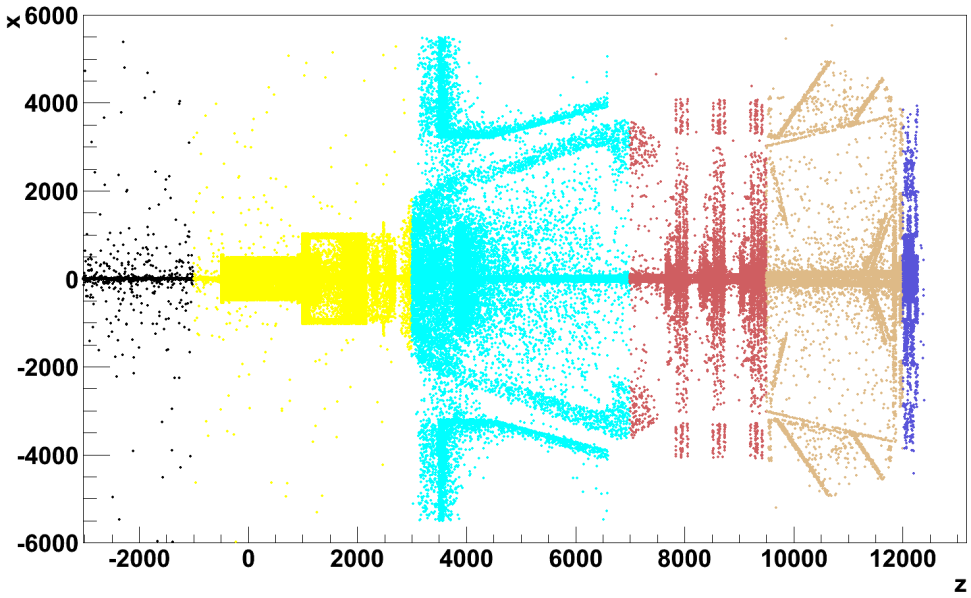

Figure 2: LHCb top view shape obtained with the simulated position of material interactions

Acknowledgments

There are plenty of people who have helped me in very different ways all along these five years for the completion of my PhD. And the fact that I have been lucky enough to be part of a collaboration such as LHCb makes this even more important in my case. My gratitude to all of them and my apologies to all of those I do not explicitly mention in here.

I would like to begin expressing my appreciation to my supervisors, Jose Ángel Hernando and Bernardo Adeva, for their support and time dedicated to my learning. They have a big responsibility in turning me into a Particle Physicist.

Given our close work together, I specially thank Diego Martínez for his help and friendship these 5 years (and for many hours sitting together in front of a computer solving issues in a very wide range of subjects). I would also like to thank Frederic Teubert for his help from the very beginning, in my Summer Student times. My special appreciation also to Marco Bettler, who broke to me many stereotypes about Swiss and with whom I have shared a lot of work in $B_s^0 \to \mu^+\mu^-$ and $K_S^0 \to \mu^+\mu^-$.

I would also like to show my special gratitude to Pierluigi Campana, Carlos Pajares, Thomas Ruf, Cibrán Santamarina and Frederic Teubert for kindly accepting being members of my thesis jury.

Since I have been involved in several working groups at LHCb, there is different people that I would like to mention as well in here. First, Hans Dijkstra for trusting my HLT work me when I was really inexpert. In the MuonID side, Erica Polycarpo, with who I enjoyed working and had a wonderful time in Rio (and who I also thank for having invited me to go!), Helder Lopes, Gaia Lanfranchi, Matteo Palutan and Barbara Sciascia. Thanks to all my colleagues in the $B_s^0 \to \mu^+\mu^-$ working group, and in particular Gaia, Matteo and Barbara (again), Johannes Albrecht, Hugo Ruiz, Laurence Carson, Cosme Adrover, Fatima Soomro, Flavio Archilli and Francesco Dettori. I would also like to show my acknowledgment to Thomas Ruf, with whom I worked in several different periods of my PhD and in particular during last months, while preparing $K_S^0 \to \mu^+\mu^-$ analysis. Finally, my gratitude to the LHCb catalan gang: Marc, Richi, Antonio, Albert, Elías, Xavi and Miriam.

I thank the financial support provided by the "Programa de Formación de Profesorado Universitario" of the "Ministerio de Educación, Cultura y Deporte", and by the "Consellería de Educación" during the first months of my thesis.

Gustaríame ter tamén unha palabra de agradecemento con toda a xente coa que tiven o pracer de traballar na Universidade de Santiago. Téñome sentido afortunado moitas veces de atopar un grupo de xente tan boa, moitos dos cales xa considero amigos! Quixera mencionar aos meus compañeiros de fatigas como estudantes de doutoramento: Diego, Lucas, Celestino, Paula, Álvaro, e Brais; e ao resto do grupo co cal, dun xeito ou doutro, ou traballei ou pasei tempo: Abraham, Cibrán, Pablo, Eliseo, Marcos, Laurence, Juan, Toño, Víctor, Daniel, Antonio, Máximo e Pablo.

Xa desde un punto de vista máis persoal, grazas a moitos amigos cos que xa perdín contacto, seguramente en parte polo tempo adicado a esta tese, e polo cal pido desculpas. En particular quixera mencionar a algúns dos que aínda teño a sorte de manter: Álex, Noelia e Bea, aos que coñecín hai xa 10 anos, sendo case outra persoa, pero que aínda así tiveron a paciencia de seguirme aturando desde aquela!

Grazas a meus pais, Ángeles e Ramón, e a miña avoa Peregrina. Polo seu amor, axuda e apoio constantes, sen os cales nin tería chegado a esta tese nin sería a persoa que hoxendía son.

Finalmente, quixera darlle as grazas a Nery. Por complementarme, ter paciencia, saber escoitarme e estar sempre ao meu lado. Por regalarme un sorriso nos momentos nos que máis o preciso. E porque sen ela tampouco nunca tería chegado aquí.

Everyone who has ever taken a shower has had an idea. It's the person who gets out of the shower, dries off and does something about it who makes a difference.
Nolan Bushnell

Science is the belief in the ignorance of experts.
Richard Feynman

Abstract

This thesis is devoted to the explanation of the searches for the very rare decays $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$, performed using the 1 fb^{-1} of data taken by LHCb during 2011 from the LHC proton–proton collisions at an energy of $\sqrt{s}$=7 TeV.

The most relevant theory aspects concerning these searches will be presented. Furthermore, the LHCb experiment will be described, and some of its most interesting results in the rare decays area reported.

The $B^0_{(s)} \to \mu^+\mu^-$ search has not resulted in any signal excess above background, but has produced world best upper limits in the branching ratios of both channels: $\mathcal{B}(B^0 \to \mu^+\mu^-) < 1.0 \times 10^{-9}$ and $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 4.5 \times 10^{-9}$, both at 95% confidence level.

Finally, no significant signal excess has been found in the search for $K^0_S \to \mu^+\mu^-$ neither, and again an upper limit on its branching ratio has been set: $\mathcal{B}(K^0_S \to \mu^+\mu^-) < 11.2 \times 10^{-9}$ at 95% confidence level. This upper limit is a factor of 35 better than the previous world best.

Limiar

Esta tese está adicada á explicación das procuras dos decaementos moi raros $B^0_{(s)} \to \mu^+\mu^-$ e $K^0_S \to \mu^+\mu^-$, levadadas a cabo empregando o fb^{-1} de datos recollidos por LHCb durante o ano 2011 a partires das colisións protón–protón do LHC a unha enerxía de $\sqrt{s}$=7 TeV.

Presentaranse os aspectos teóricos máis relevantes para estas procuras. Ademais, describirase o experimento LHCb, e relataranse algúns dos resultados máis interesantes na área dos decaementos raros.

A busca de $B^0_{(s)} \to \mu^+\mu^-$ non ten resultado en ningún exceso significativo de sinal sobre o fondo, pero ten producido os mellores límites superiores mundiais nas fraccións de desintegración das dúas canles: $\mathcal{B}(B^0 \to \mu^+\mu^-) < 1.0 \times 10^{-9}$ e $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 4.5 \times 10^{-9}$, as dúas cun nivel de confianza do 95 %.

Finalmente, tampouco se ten atopado ningún exceso de sinal na procura de $K^0_S \to \mu^+\mu^-$, e novamente téñense posto límites superiores na súa fracción de desintegración: $\mathcal{B}(K^0_S \to \mu^+\mu^-) < 11.2 \times 10^{-9}$ ao 95% de nivel de cofianza. Este límite superior é un factor 35 mellor ca o anterior mellor mundial.

Contents

1 Introduction

The Standard Model (SM) of Particle Physics is our most successful theory to explain the structure and relations[1] of all the particles which compose the matter around us. The SM establishes three fundamental interactions, weak, strong and electromagnetic, with which it gives an explanation to most of the phenomena observed in Nature. There are, however, certain measurements that cannot be explained by the SM (or, more generically, provoke "tensions" to it). The most typical example of these is the evidence for Dark Matter. Other examples are the oscillation of neutrinos, the matter/anti-matter unbalance or the fine-tuning problems. Because of this, New Physics (NP) models have been developed in the last years to cope with these problems, going beyond the SM. The advantage of the NP models is that they yield predictions that can be tested in different High Energy Physics (HEP) experiments.

The largest and most important experiment to test the SM and NP models is the Large Hadron Collider (LHC), located at CERN, in Geneva (Switzerland). The LHC collides protons almost head–to–head at a nominal center of mass energy of $\sqrt{s} = 14$ TeV ($\sqrt{s} = 7$ TeV during 2011), which are profited by four different detectors (ATLAS, CMS, LHCb and ALICE) to perform experimental HEP searches with a never before achieved precision. Even if, as just said, ALICE is also recording proton–proton collisions, it should be noted that it is mainly devoted to the research of a new state of matter called Quark–Gluon Plasma, for which it needs the LHC to collide heavy ions in special runs.

In particular, the work of this thesis has been done within the LHCb collaboration. As just explained, LHCb is one of the four big LHC experiments at CERN. LHCb is specialised in CP–violation measurements and in the search for rare decays, mainly by the study of B mesons. With this purpose, the LHCb detector has been designed as a forward spectrometer, recording the low polar angle products of the proton–proton collisions. Moreover, and again given its physics goals, LHCb requires excellent vertexing, impact parameter, momentum and mass resolutions; very discriminating particle identification and a flexible trigger. The results to be presented here are based on the 1 fb^{-1} of data taken by LHCb during year 2011.

Some of the most remarkable LHCb results so far come in the rare decays area. The rare decays analyses include several common experimental features, what makes easy to share tools and experience on solving issues among all of them. Some of the key rare decays for LHCb are the search for $B^0_{(s)} \to \mu^+\mu^-$, the analysis of the angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ and the study of radiative decays such as $B^0 \to K^{*0}\gamma$. $K^0_S \to \mu^+\mu^-$ is also a rare decay whose search has been proposed as a novel way to profit the excellent performance of the LHCb detector.

As already said, the search for $B^0_{(s)} \to \mu^+\mu^-$ is one of the most relevant analysis for LHCb, since both the branching ratios (hereinafter referred to as $\mathcal{B}$ or BR) $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ are very interesting from a theoretical point of view. While very suppressed and well determined in the SM, these decays can be enhanced (or even more suppressed) by several NP models, so that they are very useful in order to constraint NP scenarios. One of the objectives of this thesis is, therefore, either the measurement of the $\mathcal{B}$ of these modes or the setting of upper limits on it. In

[1] except for the gravitation.

order to do so, a strategy has been designed with the idea of reducing the dependence on Monte Carlo simulated events, basing all the parameter calibration on data. Furthermore, the obtention of the $\mathcal{B}$ is achieved by normalising to other channels with similar geometry and trigger, such as $B^+ \to J/\psi K^+$.

The last goal of this thesis is the study of the very rare decay $K^0_S \to \mu^+\mu^-$. This channel shares some of the theoretical and experimental characteristics of $B^0_{(s)} \to \mu^+\mu^-$, but with some remarkable differences. Its $\mathcal{B}$ is also sensitive to NP, although it is in principle much more suppressed than that of both $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$. From the experimental point of view, the B^0 and B^0_s masses are a factor of more than 10 larger than that of K^0_S, while they both fly a factor of more than 50 less, which implies very different reconstruction and trigger performances. The search for $K^0_S \to \mu^+\mu^-$ also uses data for the calibration of the most relevant parameters, with the $\mathcal{B}$ being obtained by normalising to the $K^0_S \to \pi^+\pi^-$ decay.

During my PhD period at LHCb, I have worked in the hadron trigger, in the muon identification and in three different analysis (the searches for $B^0_{(s)} \to \mu^+\mu^-$, $B^0_s \to K^{*0}\bar{K}^{*0}$ and $K^0_S \to \mu^+\mu^-$). As explained, among all these just the work in the $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$ is included in this thesis.

My work in the $B^0_{(s)} \to \mu^+\mu^-$ analysis began with several technical aspects, such as the ntuple creation and the preparation of Monte Carlo simulated data with the right trigger configuration. From the physics point of view, I had responsibility in the tuning of the selection variables and in the calibration of the Multi-Variate-Analysis tool both in signal and background. I also worked in the study of the background originated by $B^0_{(s)} \to h^+h'^-$ in which both hadrons are misidentified as muons. Finally, I also played an important role in the search for specific backgrounds.

As far as the $K^0_S \to \mu^+\mu^-$ analysis is concerned, I am responsible for proposing it as a novel way to profit the excellent LHCb performance, extending the collaboration physics programme to the kaon sector. Together with two colleagues, I designed the LHCb $K^0_S \to \mu^+\mu^-$ analysis and worked almost in every aspect of it. Examples are the creation of ntuples, development of the selection, calibration and tuning of the Multi-Variate-Analysis tool, study of specific backgrounds, normalisation and determination of the expected background.

This thesis is divided as follows. Chapter 2 overviews the most important theory aspects for the search of both $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$. These include a brief explanation of the SM and its predictions on $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$ and $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ and a justification about the need of NP, also stating how the searched decays can affect the NP scenarios. The LHCb experiment at the LHC is presented in chapter 3, describing the detector and how the results are achieved, starting from raw data from proton–proton collisions. The performance of LHCb during 2011 is also briefly reported. A summary on the most important 2011 rare decays analyses by LHCb will be given in chapter 4. Chapter 5 explains in detail the search for $B^0_{(s)} \to \mu^+\mu^-$ at LHCb, and it includes two brief appendices to report on the theory implications of the result and on the combination of this result with those from other LHC experiments. The $K^0_S \to \mu^+\mu^-$ analysis is extensively presented in chapter 6. Finally, conclusions are drawn in chapter 7 and a summary of the thesis can be found in chapter 8 in English and in chapter 9 in Galician.

2

Theory overview

2.1 Standard Model

2.1.1 Introduction

The Standard Model (SM) is a quantum field theory (QFT) describing strong and electroweak (EW) interactions. The SM structure is explained next, following essentially [1]. The strong interactions are described by Quantum Chromodynamics (QCD) corresponding to the symmetry group $SU(3)_C$ of color (C), while the EW interaction is described by the group $SU(2)_T \otimes U(1)_Y$ of weak-isospin (T) and hypercharge (Y), being then $SU(3)_C \otimes SU(2)_T \otimes U(1)_Y$ the full group of gauge symmetry for the SM.

$$G_{SM} = SU(3)_C \otimes SU(2)_T \otimes U(1)_Y \tag{2.1}$$

This symmetry is spontaneously broken into $SU(3)_C \otimes U(1)_{EM}$ by the vacuum expectation value (VEV) of (the neutral component of) a scalar isospin doublet, with hypercharge 1/2, called Higgs:

$$G_{SM} \overset{Higgs(1,2)_{1/2}}{\rightarrow} SU(3)_C \otimes U(1)_{EM}$$

As a result of the interaction with the Higgs field, EW bosons combine into the massive particles $W^{\pm}$ and Z^0 and the massless photon. The interaction with Higgs gives also masses to the fermions. Each fermion generation, out of a total of three, has five representations of the SM gauge symmetry:

$$Q_{L,i}(3,2)_{+1/6} \qquad U_{R,i}(3,1)_{+2/3} \qquad D_{R,i}(3,1)_{-1/3} \qquad L_{L,i}(1,2)_{-1/2} \qquad E_{R,i}(1,1)_{-1}$$

The subscript number is the hypercharge, and the numbers in parenthesis indicate if it acts as a triplet or singlet in $SU(3)_C$ and as a doublet or singlet in $SU(2)_T$. The subscript i = 1,2,3 indicates the fermion generation, and the L,R, the chirality.

The EW symmetry breaking (EWSB) and the effects induced by Higgs field such like CP violation and flavour depending processes are explained in section 2.1.2.2. The fermion and boson content of SM is explained in more detail in sections 2.1.2.3 and 2.1.2.4.

Thus, the SM Lagrangian can be decomposed in three parts:

$$\mathcal{L} = \mathcal{L}_{Kin} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yuk} \tag{2.2}$$

where the kinetic part includes the corresponding covariant derivative to preserve the gauge invariance, the Higgs part includes Higgs self interactions and Yukawa part includes Higgs-fermion interactions.

2.1.2 Mass generation and eigenstates

2.1.2.1 Boson masses and EWSB

A Lagrangian containing only the terms of the gauge symmetry is not enough to build a model where the particles are massive. The gauge bosons are massless if the symmetry is unbroken, and masses for the fermions as self-interactions such like $\overline{\Psi}_L\Psi_R$ (Dirac mass) or $\Psi_L\Psi_R$ (Majorana mass) would explicitly break the SU(2) symmetry. Non-abelian broken gauge theories are not renormalisable, thus in the SM the masses of the EW gauge bosons and the fermions are given by a spontaneously symmetry breakdown. This is achieved by the introduction of the Higgs, a scalar isospin doublet with hypercharge +1/2:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

This doublet has a self interaction of the form:

$$\mathcal{L}_{Higgs} = \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \tag{2.3}$$

The first term is similar to a mass one, but with opposite sign. Such quadratic potential does not minimise at 0 (see example in figure 2.1), and thus acquires a VEV $\nu = \mu/\sqrt{\lambda}$.

$$\langle\phi\rangle_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu \end{pmatrix} \tag{2.4}$$

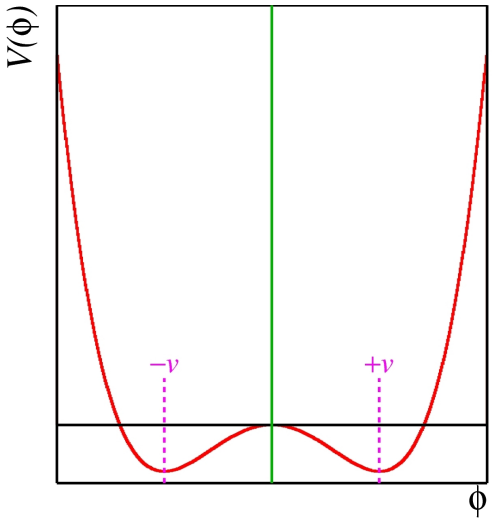

Figure 2.1: Higgs potential. The minimum is not at 0, and therefore the potential has a VEV. Figure taken from [2]

The VEV gives masses, through the Higgs kinetic term plus the Higgs self-interaction Lagrangian, to the following boson combinations:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^{(1)} \mp W_\mu^{(2)}) \rightarrow M_W = g \cdot \frac{\nu}{2}$$

$$Z_\mu^0 = \frac{1}{\sqrt{g^2+g'^2}}(gW_\mu^{(3)} - g'B_\mu^{(2)}) \rightarrow M_Z = \sqrt{g^2+g'^2} \cdot \frac{\nu}{2}$$

From the degrees of freedom of the original Higgs field:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} G_1^+ + iG_2^+ \\ \nu + (H^0 + iG_3^0) \end{bmatrix}$$

H^0 will be a massive scalar particle, having the massless Goldstone bosons G_i "eaten" by the gauge bosons $W^\pm$ and Z^0, giving rise to their longitudinal polarisations and masses.

2.1.2.2 Fermion masses and CKM matrix

In order to give masses to the fermions, the corresponding couplings between them and the Higgs field are added, while keeping the Lagrangian SU(2) invariant. For example, for a single generation:

$$\Delta\mathcal{L} = -\lambda_e \bar{E}_L \phi E_R - \lambda \bar{Q}_L \phi D_R - \lambda_u \epsilon^{ab} \bar{Q}_{La} \phi_b^\dagger U_R + h.c. \tag{2.5}$$

Substituting the VEV the fermion masses have the form:

$$m_e = \frac{\nu \cdot \lambda_e}{\sqrt{2}} \qquad m_u = \frac{\nu \cdot \lambda_u}{\sqrt{2}} \qquad m_d = \frac{\nu \cdot \lambda_d}{\sqrt{2}}$$

These λ_i are inputs to the SM and thus allow having very different masses for different fermions. When the three fermion generations are added to the theory, additional terms mixing quarks of different generations are possible. Alternatively, it is possible to diagonalise the Higgs couplings by switching to a different basis for the quark fields. Writing the Lagrangian in this alternative basis (hereinafter referred to as "mass basis" or "physical basis") will of course simplify $\mathcal{L}_{Yuk}$ but with the cost of causing a complication in the gauge side. Calling q the interaction eigenstates and q' the mass eigenstates, both bases are related through the unitary relations:

$$u_L^i = U_u^{ij} u_L'^j \qquad d_L^i = U_d^{ij} d_L'^j$$

And thus the weak current $\bar{u}_L^i \gamma^\mu d_L^i$ transforms to $\bar{u'}_L^i \gamma^\mu (U_u^\dagger U_d)_{ij} d_L'^i \equiv \bar{u'}_L^i \gamma^\mu V_{ij}^{CKM} d_L'^i$. Being V^{CKM} called the CKM matrix [3, 4] (from Cabibbo-Kobayashi-Maskawa). Its coefficients are usually written as:

$$V^{CMK} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.6}$$

V^{CKM} is not diagonal (the experimental value of the coefficients can be found in [5]) and such structure allows transitions between the different quark generations, giving rise to processes in which quarks change flavour without changing its electric charge. These processes are called Flavour Changing Neutral Currents (FCNC) and in particular include the decays $B_s^0 \to \mu^+\mu^-$ and $K_S^0 \to \mu^+\mu^-$. CP violation also arises from the non diagonal structure of V^{CKM} requiring, in addition, the presence of three different generations. Equivalently, if V^{CKM} were the identity matrix CP violation and FCNCs would not exist within the SM.

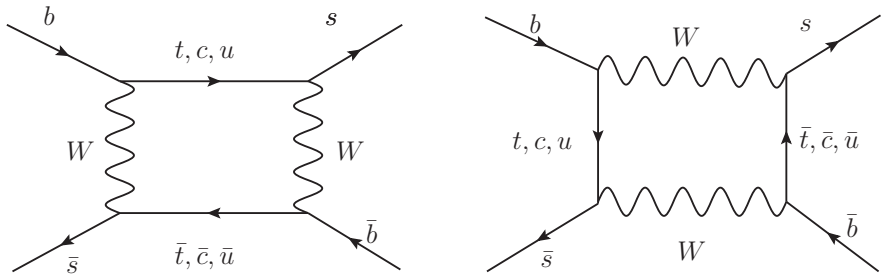

Figure 2.2: Diagrams contributing to B_s^0 oscillation

Another particular but very important example of process arising from the fact that V^{CKM} is different from the identity matrix is the oscillation of neutral mesons composed by quarks of different generations. The off-diagonal terms of CKM matrix allow particles such like D^0, K^0, B^0 or B_s to perform particle-antiparticle oscillations (see figure 2.2 for examples of diagrams involving B_s oscillation).

The number of free parameters of the CKM matrix with the three generations is 4: 3 rotation angles and one phase that is the only source of CPV in the SM. It can be noticed that in the case of only two generations that phase can be removed, which implies that CPV processes must involve the three quark families.

2.1.2.3 Fermions

The Standard Model fermions can be divided in two groups, depending on whether they are affected by strong interaction (quarks) or not (leptons). Each quark has three possible color states and (at low energy) only exists in bound states of color singlets, called hadrons. Hadrons are then composed by quarks (and gluons, the gauge bosons of QCD), being the most common states quark - antiquark (mesons), and three quarks (baryons). Due to spin addition, baryons are also fermions, while mesons are bosons.

Leptons are e, μ, τ and a neutrino (ν) for each one. In the SM neutrinos are massless particles so that their helicity becomes equivalent to chirality. This means that there are not right-handed neutrinos in the SM and, equivalently, there are not left-handed antineutrinos[1].

		T	T_3	Y	Q
Leptons	$\nu_e,\ \nu_\mu, \nu_\tau$	$\frac{1}{2}$	$\frac{1}{2}$	-1	0
	$e_L,\ \mu_L, \tau_L$	$\frac{1}{2}$	$-\frac{1}{2}$	-1	-1
	$e_R,\ \mu_R, \tau_R$	0	0	-2	-1
Quarks	$u'_L,\ c'_L, t'_L$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$
	$u'_R,\ c'_R, t'_R$	0	0	$\frac{4}{3}$	$\frac{2}{3}$
	$d'_L,\ s'_L, b'_L$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{3}$
	$d'_R,\ s'_R, b'_R$	0	0	$-\frac{2}{3}$	$-\frac{1}{3}$

Table 2.1: SM fermions

2.1.2.4 Bosons

Apart from mesons, the SM contains the gauge bosons corresponding to strong and EW interactions, and Higgs (H^0) boson, responsible for the masses of SM particles.

The gauge bosons of QCD are massless particles of spin 1, called gluons, and have eight possible color states. QCD couplings have the property of becoming small at high energies (or small distances); this effect is known as "asymptotic freedom".

The gauge bosons corresponding to $SU(2)_T \otimes U(1)_Y$ are W^i_μ (i = 1,2,3) and B_μ, for SU(2) and U(1) respectively, and the four should be massless in order to conserve the gauge symmetry. However, the symmetry breaking induced by Higgs field changes them into W^+, W^-, Z^0 and photon (A^μ), where only the photon is massless. All of them have spin 1.

2.1.3 $B^0_{(s)} \to \mu^+\mu^-$ in Standard Model

The $B^0_{(s)} \to \mu^+\mu^-$ decays (where $B^0_{(s)}$ refers both to B^0 and B^0_s mesons, i.e., a combination of quark $\bar{b}$ with either d or s), are FCNC which do not arise at tree level in SM. It should be noted that the tag $B^0 \to \mu^+\mu^-$ ($B^0_s \to \mu^+\mu^-$) refers to the decay of both B^0 and $\bar{B}^0$ (B^0_s and $\bar{B}^0_s$) to the final state $\mu^+\mu^-$. The reason for these decays not taking place at tree level in the SM is that the two final state muons can only be directly produced from a photon, Higgs, or Z^0, but none of these bosons can be originated by a $b - d/s$ quark interaction directly, since they can be produced only by particles of the same flavour. This is common to all the FCNC (such as $K^0_S \to \mu^+\mu^-$). The main contributions to these decays are

[1] It is now known experimentally that neutrinos suffer flavour oscillations, which means they are not massless. As it will be later seen this can be fitted into the SM with small modifications to it.

weak interactions of fourth order: Z^0 penguins and $W^\pm$ box. The corresponding diagrams are shown in 2.3 and were first calculated in [6] for the case of $K_L^0 \to \mu^+\mu^-$. Quarks entering in the loops can be u, c and t, but due to larger values of V_{tb} and m_t, u and c contributions can be neglected with respect to t contribution. As explained before, those diagrams would be 0 for $V^{CKM} = 1$.

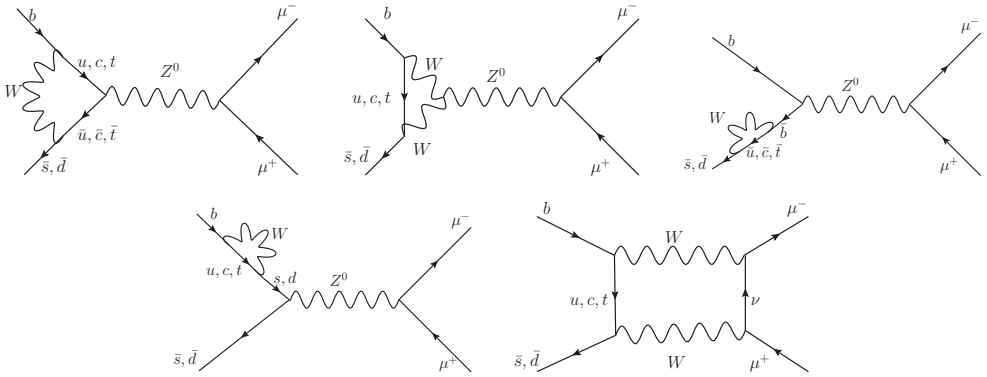

Figure 2.3: Diagrams contributing to $B^0_{(s)} \to \mu^+\mu^-$ decays in SM

2.1.3.1 Effective Hamiltonian and Wilson Coefficients

In order to analytically study the amplitude of the $B^0_{(s)} \to \mu^+\mu^-$ decays, an effective Hamiltonian is used in order to account for all the contributions in diagrams in 2.3. This happens typically in hadronic weak decays. These Hamiltonians are expressed in terms of Operator[2] Product Expansion (OPE):

$$H_{eff} = G_F \sum_i C_i \cdot Q_i$$

where G_F is the Fermi constant, Q_i are local operators including only the initial and final state fields, multiplied by Wilson [7] coefficients, C_i, that contain the information about short distance physics, such as the masses of particles entering in internal loops. An example of such effective descriptions was Fermi theory of neutron decay [8] $n \to p + e^- + \bar{\nu}_e$, where the degrees of freedom corresponding to the W boson exchanged were integrated out giving a Hamiltonian with a four fermion local interaction:

$$H_{eff} = G_F(\bar{p}\gamma^\mu n)(\bar{e}\gamma^\mu \nu_e) + h.c.$$

In the case of $B_q \to l^+l^-$ transitions (where q is the quark accompanying b), and in order to classify different contributions not only in the SM but also in its extensions, the effective Hamiltonian is often written as in [9]:

$$H_{eff} = \frac{G_F \alpha}{\sqrt{2}\pi sin^2(\theta_W)} V_{tb}^* V_{tq} \left[C_S \hat{Q}_S + C_P \hat{Q}_P + C_A \hat{Q}_A + C'_S \hat{Q}'_S + C'_P \hat{Q}'_P + C'_A \hat{Q}'_A \right] \tag{2.7}$$

where α is the electromagnetic fine structure constant, θ_W the Weinberg angle, and the operators (note different quark mass in S', P' with respect to S, P):

$$\hat{Q}_S = m_b(\bar{q}P_R b)(\bar{l}l) \qquad \hat{Q}_P = m_b(\bar{q}P_R b)(\bar{l}\gamma_5 l) \qquad \hat{Q}_A = (\bar{q}\gamma^\mu P_L b)(\bar{l}\gamma_\mu\gamma_5 l)$$
$$\hat{Q}'_S = m_q(\bar{q}P_L b)(\bar{l}l) \qquad \hat{Q}'_P = m_q(\bar{q}P_L b)(\bar{l}\gamma_5 l) \qquad \hat{Q}'_A = (\bar{q}\gamma^\mu P_R b)(\bar{l}\gamma_\mu\gamma_5 l)$$

There, m_b is the mass of the b quark, $(q = d, s)$ is the field of the companion quark of the b in the B_q meson, $P_{R,L} = (1 \pm \gamma^5)$, l is the lepton field and the subindices A, P, S refer to axial, pseudoscalar and scalar contributions, respectively. Using the Hamiltonian in equation 2.7, $\mathcal{B}(B_q \to \mu^+\mu^-)$ can be

[2] In quantum mechanics "operators" are associated with the measurable parameters in a physical system.

expressed as:

$$\mathcal{B}(B_q \to \mu^+\mu^-) = \frac{G_F^2\alpha^2}{64\pi^3 sin^4\theta_W}|V_{tb}^*V_{tq}|^2\tau_{B_q}M_{B_q}^3 f_{B_q}^2\sqrt{1-\frac{4m_\mu^2}{M_{B_q}^2}}\times$$
$$\times\left\{M_{B_q}^2\left(1-\frac{4m_\mu^2}{M_{B_q}^2}\right)\left(\frac{C_s-\mu_q C_s'}{1+\mu_q}\right)^2+\left[M_{B_q}\left(\frac{C_P-\mu_q C_P'}{1+\mu_q}\right)+\frac{2m_\mu}{M_{B_q}}(C_A-C_A')\right]^2\right\} \tag{2.8}$$

where τ_{B_q}, M_{B_q} and f_{B_q} are the mean lifetime, mass and decay constant[3] of B_q meson, m_μ the muon mass and μ_q the ratio of masses m_q/m_b. As $C_{S,P}$ and $C'_{S,P}$ are in general of comparable size [9] and $\mu_q << 1$, the terms with μ_q can be neglected simplifying the expression:

$$\mathcal{B}(B_{(s)}^0 \to \mu^+\mu^-) = \frac{G_F^2\alpha^2}{64\pi^3 sin^4\theta_W}|V_{tb}^*V_{tq}|^2\tau_{B_q}M_{B_q}^3 f_{B_q}^2\sqrt{1-\frac{4m_\mu^2}{M_{B_q}^2}}\times$$
$$\times\left\{M_{B_q}^2\left(1-\frac{4m_\mu^2}{M_{B_q}^2}\right)C_S^2+\left[M_{B_q}C_P+\frac{2m_\mu}{M_{B_q}}(C_A-C_A')\right]^2\right\} \tag{2.9}$$

It must be said that, using same dimensions for the different Wilson coefficients, the axial contributions are suppressed with respect to scalar and pseudoscalar by a factor of $\sim \frac{m_\mu}{M_{B_q}}$ (helicity suppression). Moreover, the axial contributions are already constrained by other measurements in $b \to sl^+l^-$ transitions, such as $B^0 \to K^{*0}\mu^+\mu^-$, which has shown to be more sensitive to axial than to scalar contributions [11]. Because of all of this, $\mathcal{B}(B_{(s)}^0 \to \mu^+\mu^-)$ is expected to be more sensitive to new physics in the scalar sector.

Even if in LHCb the number of produced B^0 mesons is around four times larger than the number of B_s (this given by the fraction f_d/f_s [12]), the main $B_{(s)}^0 \to \mu^+\mu^-$ decay to search for, in case the $\mathcal{B}$s obtained within the SM were the correct, would be the B_s, since the B^0 mode would be suppressed by $\sim \frac{|V_{tb}^*V_{td}|^2}{|V_{tb}^*V_{ts}|^2} \sim \frac{1}{20}$.

2.1.3.2 QCD corrections

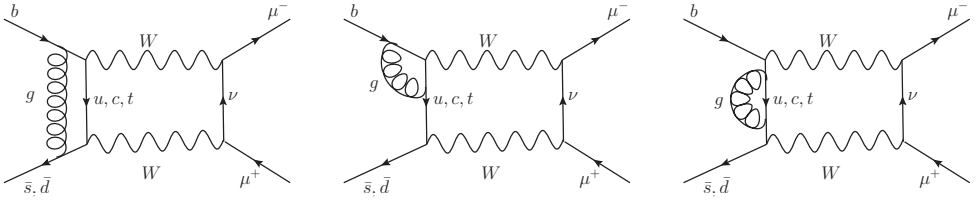

Figure 2.4: Examples of QCD diagrams contributing to $B_{(s)}^0 \to \mu^+\mu^-$ decays in SM

Gluon exchange between the quarks b, $s(d)$ and u, c, t entering in the diagrams (see for instance those in 2.4) needs to be included in order to get the right value of Wilson coefficients. Furthermore, the relation of $\mathcal{B}(B_{(s)}^0 \to \mu^+\mu^-)$ to the B_q oscillation frequency, pointed out in [13], allows to reduce the uncertainty, yielding [14]:

$$\boxed{\mathcal{B}(B^0 \to \mu^+\mu^-)|^{SM} = (1.1 \pm 0.1) \times 10^{-10}} \tag{2.10}$$

$$\boxed{\mathcal{B}(B_s^0 \to \mu^+\mu^-)|^{SM} = (3.2 \pm 0.3) \times 10^{-9}} \tag{2.11}$$

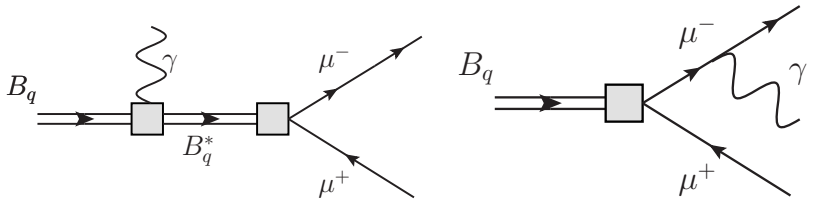

Figure 2.5: Examples of diagrams contributing to $B^0_{(s)} \to \mu^+\mu^-\gamma$ decays in SM. Left diagram shows initial state radiation (ISR), while right diagram final state radiation (FSR). The FSR diagrams are composed by the bremsstrahlung of the final state muons. The small boxes represent all the boxes and penguins seen in figure 2.3.

2.1.3.3 QED corrections

The $B^0_{(s)} \to \mu^+\mu^-\gamma$ decays are closely related to $B^0_{(s)} \to \mu^+\mu^-$, but depending on the origin of the photon they should be considered as a "component" of it or as an independent channel. Indeed, if the photon arises as an initial state radiation (ISR), the decay is independent. However, if the photon appears as a bremsstrahlung of the final state muons (final state radiation or FSR), then $B^0_{(s)} \to \mu^+\mu^-\gamma$ is a part of $B^0_{(s)} \to \mu^+\mu^-$. The diagrams in figure 2.5 show examples of both FSR and ISR. Since the amplitude includes both the ISR and FSR, the final $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-\gamma)$ will have contributions of ISR, FSR and the interference between both.

It must be also said that the computation of $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-\gamma)$ diverges unless a lower cut in the energy of the photons is imposed. This is crucial for the experimental consideration of $B^0_{(s)} \to \mu^+\mu^-\gamma$ as a background for $B^0_{(s)} \to \mu^+\mu^-$, since only the cases with soft photons can survive the cuts in the mass imposed at "selection level".

The studies in [15], in which the cut in the lower energy of the photon is set at 20 MeV, yield total $\mathcal{B}(B^0 \to \mu^+\mu^-\gamma) = 1.34 \times 10^{-10}$ and $\mathcal{B}(B^0_s \to \mu^+\mu^-\gamma) = 1.89 \times 10^{-8}$. It is important to remark that this BR includes the resonant contributions, such as $B^0_s \to \phi(\mu^+\mu^-)\gamma$. Moreover, it includes both the FSR and ISR. The dimuon q^2 spectra estimated in [15] are shown in figure 2.6, for the cases of B^0 and B^0_s. The fraction of the total spectra around the B^0 and B^0_s mass is seen to be rather small. A complete MC simulation has also been done following the model in [15]. This will be used in section 5.6.4.3 to estimate the possible contribution of $B^0_s \to \mu^+\mu^-\gamma$ to the calculated $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$, according to this model (the contribution of $B^0 \to \mu^+\mu^-\gamma$ is expected to be two orders of magnitude smaller).

Finally, the possible contribution of $B^0_s \to \mu^+\mu^-\gamma$ to the measured $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ has been studied in [16, 17]. This latest study, which focuses on the soft photons (which could affect more the measured $\mathcal{B}[B^0_s \to \mu^+\mu^-]$), preliminarily suggests that a non-negligible fraction of the BR measured by LHCb, up to a 8%, may correspond to soft radiation from $B^0_s \to \mu^+\mu^-\gamma$. The study performed in [14] also produces similar conclusions.

2.1.3.4 Corrections because of $\Delta\Gamma_s \neq 0$

One last correction that should be applied in order to compare the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ predicted theoretically to the one obtained experimentally has been introduced in [18, 19]. This correction comes from the fact that, in contrast to the B^0 system, the B^0_s mesons exhibit a sizeable difference between the decay widths of the light and heavy mass eigenstates, $\Gamma^{(s)}_L$ and $\Gamma^{(s)}_H$, respectively. The most precise measurement of this difference was given by the LHCb collaboration [20], from the time-dependent analysis for $B^0_s \to J/\psi\,\phi$.

$$y_s \equiv \frac{\Delta\Gamma_s}{2\Gamma_s} = \frac{\Gamma^{(s)}_L - \Gamma^{(s)}_H}{2\,\Gamma_s} = 0.088 \pm 0.014,$$

where

$$\tau^{-1}_{B^0_s} \equiv \Gamma_s \equiv \frac{\Gamma^{(s)}_L - \Gamma^{(s)}_H}{2} = (0.6580 \pm 0.0085)\ \text{ps}^{-1}$$

[3] the decay constant, f_P, of a pseudoscalar meson P (such as B^0_s and B^0) with 4-momentum q is defined as [10] $< P(q)|\bar{q}\gamma_\mu\gamma_5 q'|0 > \equiv -if_P q_\mu$

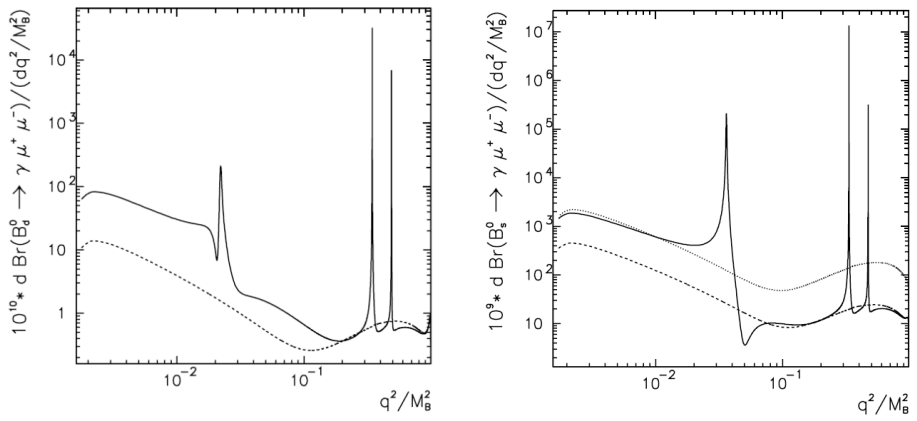

Figure 2.6: The dimuon q^2 spectrum estimated in [15] for the case of $B^0 \to \mu^+\mu^-\gamma$ (left) and $B_s^0 \to \mu^+\mu^-\gamma$ (right). The solid line corresponds to a cut in the lower energy of the photon at 20 MeV. The other lines, correspond to cuts at higher energies. The q^2 is shown in units of the $M(B^0)^2$ and $M(B_s^0)^2$ mass.

In practice, the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is determined experimentally from the total event yield, ignoring information on the lifetime of the particles. With this,

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{exp} \equiv \frac{1}{2}\int_0^\infty \left\langle \Gamma(B_s^0(t) \to \mu^+\mu^-)\right\rangle dt \tag{2.12}$$

where the total untagged rate is given by:

$$\begin{aligned}\left\langle \Gamma(B_s^0(t) \to \mu^+\mu^-)\right\rangle &= \Gamma(B_s^0(t) \to \mu^+\mu^-) + \Gamma(\bar{B}_s^0(t) \to \mu^+\mu^-) \\ &\propto e^{-\frac{t}{\tau_{B_s^0}}}\left[\cosh\left(y_s\frac{t}{\tau_{B_s^0}}\right) + \mathcal{A}_{\Delta\Gamma}\sinh\left(y_s\frac{t}{\tau_{B_s^0}}\right)\right]\end{aligned}$$

$\mathcal{A}_{\Delta\Gamma}$ being an unknown model dependent parameter in the range $[-1,+1]$ ($\mathcal{A}_{\Delta\Gamma}^{SM} = 1$).

Note that the quantity in equation 2.12 is essentially the average of the branching ratios for the heavy and light mass eigenstates.

On the other hand, what is generally calculated theoretically are CP-averaged decay rates in the flavour eigenstate basis, i.e.,

$$\left\langle \Gamma(B_s^0(t) \to \mu^+\mu^-)\right\rangle|_{t=0} = \Gamma(B_s^0 \to \mu^+\mu^-) + \Gamma(\bar{B}_s^0 \to \mu^+\mu^-)$$

so that:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{theo} = \frac{\tau_{B_s^0}}{2}\left\langle \Gamma(B_s^0(t) \to \mu^+\mu^-)\right\rangle|_{t=0} \tag{2.13}$$

From equations 2.12 and 2.13, it can be shown that:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{theo} = \frac{1-y_s^2}{1+\mathcal{A}_{\Delta\Gamma}y_s}\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{exp} \tag{2.14}$$

which gives a model dependent (through $\mathcal{A}_{\Delta\Gamma}$) correction to the theory prediction of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$, in order to compare to the experimental result. In the case of SM, using $\mathcal{A}_{\Delta\Gamma}^{SM} = 1$ and the value of y_s given by LHCb (0.088 ± 0.014), the results in 2.11 are corrected to:

$$\boxed{\mathcal{B}(B_s^0 \to \mu^+\mu^-)|_{y_s}^{SM} = (3.5 \pm 0.3) \times 10^{-9}} \tag{2.15}$$

Note that this correction is negligible in the case of $B^0 \to \mu^+\mu^-$, since in this case $y_s \sim 0$, which implies in equation 2.14 $\mathcal{B}(B^0 \to \mu^+\mu^-)_{theo} \sim \mathcal{B}(B^0 \to \mu^+\mu^-)_{exp}$

2.1.4 $K^0_S \to \mu^+\mu^-$ in Standard Model

Similarly to the $B^0_{(s)} \to \mu^+\mu^-$ decays, $K^0_S \to \mu^+\mu^-$ is a FCNC which does not arise at tree level in SM [21, 22]. Note that the s in $B^0_{(s)} \to \mu^+\mu^-$ refers to the quark accompanying the b in the B^0_s meson, while in K^0_S stands for "short", corresponding to one of the two mass eigenstates of the neutral kaon system (K^0_S or K^0_L). Two different components enter in the amplitude of $K^0_S \to \mu^+\mu^-$: one includes all the long-distance (LD) contributions, while the other all the short-distance (SD)[4]. While SD contributions are completely dominant in the case of $B^0_{(s)} \to \mu^+\mu^-$, the proportion between them is model dependent for $K^0_S \to \mu^+\mu^-$. In this way, LD contributions are dominant in SM. Figure 2.7 shows examples of diagrams contributing to the SD and LD components. The LD contributions are driven mainly by two intermediate virtual photons. As for the SD contributions, they are mainly composed by weak interactions of fourth order: Z^0 penguins and $W^\pm$ box.

The most general amplitude for the $K^0 \to l^+l^-$ (l being a lepton) processes includes a s-wave (A) and p-wave (B) component [23]:

$$\mathcal{A}(K^0 \to l^+l^-) = \bar{u}_l(iB + A\gamma_5)v_l \tag{2.16}$$

which yields as corresponding decay rate:

$$\Gamma(K^0 \to l^+l^-) = \frac{m_K\beta_l}{8\pi}\left(|A|^2 + \beta_l^2|B|^2\right) \tag{2.17}$$

where $\beta_l = \sqrt{1 - \frac{4m_l^2}{m_K^2}}$, m_K is the K^0 mass and m_l the final state lepton mass.

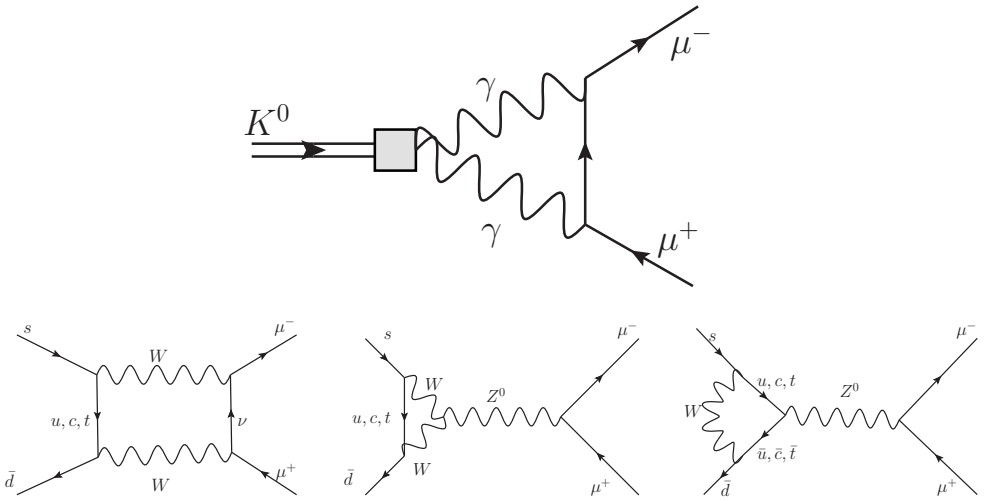

Figure 2.7: Examples of diagrams contributing to $K^0 \to \mu^+\mu^-$. Top, long distance contribution, generated by two intermediate photons. Bottom, short distance, with penguins and boxes similar to those of $B^0_{(s)} \to \mu^+\mu^-$.

These two amplitudes have opposite CP, so this constraints the way they can contribute to the K^0_S and K^0_L decay rates. In this way, CP-conserving contributions to K^0_L and K^0_S decays are generated by A and B, respectively. Because of this, the LD contributions generated by intermediate two-photon states will essentially lead to neglecting the B amplitude in the case of $K^0_L \to \mu^+\mu^-$. In opposition, and looking at the SD SM like contributions, all of them will only enter in the A term. However, in this case the CP-violating phases are expected to be rather large, so that the SD part of A has to be taken into account both in K^0_S and K^0_L decays. With all this, for $K^0_S \to \mu^+\mu^-$ equation 2.17 becomes:

$$\Gamma(K_s \to l^+l^-) = \frac{m_K\beta_l}{8\pi}\left[(\mathrm{Im}A_{SD})^2 + \beta_l^2(\mathrm{Re}B_{\gamma\gamma})^2 + \beta_l^2(\mathrm{Im}B_{\gamma\gamma})^2\right] \tag{2.18}$$

[4]The long-distance scales correspond to masses below that of the c quark, while short-distance scales correspond to masses above than (or equal to) that of the c quark.

where A_{SD} is the SD component of the A amplitude and $B_{\gamma\gamma}$ the B amplitude (LD), dominated by diagrams with two intermediate virtual photons.

2.1.4.1 Long distance contribution

Following [21], the $K^0_S \to \mu^+\mu^-$ LD decay arises mainly by two-loop diagrams of the type $K^0_S \to \gamma\gamma \to \mu^+\mu^-$ (such as those seen in figure 2.7). With these, and from [21] again, the LD contribution to the rate (terms with $B_{\gamma\gamma}$ in equation 2.18) can be written as:

$$\Gamma(K^0_S \to l^+l^-)_{LD} = \Gamma(K^0_S \to \gamma\gamma)\left(\frac{\alpha^2\beta_l^2 m_l^2}{2\pi^2|H(0)|^2 m_K^2}\,|I_{l,disp} + iI_{l,abs}|^2\right) \tag{2.19}$$

where $H(0) = 0.67$, is a function coming the one-loop amplitude calculations, α the fine-structure constant and $I_{l,disp}$ ($I_{l,abs}$) indicate the dispersive (absorptive) parts of the two-loop diagrams. For $l = \mu$, $I_{\mu,disp} \sim -2.8$ and $I_{\mu,abs} \sim 1.2$, from which:

$$\mathcal{B}(K^0_S \to \mu^+\mu^-)|_{LD} = 5.1 \times 10^{-12} \tag{2.20}$$

with an error at the level of 30%, according to [22].

2.1.4.2 Short distance contribution

The SD contributions in equation 2.18 show up in the A_{SD} term. Within SM, from [22]:

$$\mathrm{Im}A^{SM}_{short} = -\frac{G_F\alpha(M_Z)}{\pi\sin^2(\theta_W)}\sqrt{2}m_\mu F_K \mathrm{Im}(V^*_{ts}V_{td})Y(x_t) \tag{2.21}$$

where G_F is the Fermi constant, θ_W is the Weinberg angle, $\alpha(M_Z)$ is the QED coupling, F_K the Kaon form factor, V^*_{ts} and V_{td} the elements of the CKM matrix and the function $Y(x_t)$ is given in [24]. This leads to:

$$\mathcal{B}(K^0_S \to \mu^+\mu^-)|^{SM}_{SD} = 10^{-5} \times |\mathrm{Im}(\mathrm{V^*_{ts}V_{td}})|^2 = 1.4 \times 1.4 \times 10^{-12} \times \left|\frac{\mathrm{V_{cb}}}{0.041}\right|^4 \times \left|\frac{\lambda}{0.223}\right|^2 \times \bar{\eta}^2 \tag{2.22}$$

where V_{cb} is again the element of the CKM matrix, and λ and $\bar{\eta}$ are also related to the CKM matrix elements through the Wolfenstein parameterisation [25]. Given the current value of $\bar{\eta}$ from the global CKM fits [5], $\mathcal{B}(K^0_S \to \mu^+\mu^-)^{SM}_{SD}| \sim 10^{-13}$, which is smaller than the estimated error in the LD component.

This yields a SM $\mathcal{B}$ completely dominated by LD contributions, so that, using equation 2.20, one gets:

$$\boxed{\mathcal{B}(K^0_S \to \mu^+\mu^-)|^{SM} = (5.1 \pm 1.5) \times 10^{-12}} \tag{2.23}$$

2.2 New Physics

This section is devoted to the description of New Physics (NP) models which go beyond the SM, trying to improve it in those aspects in which it does not work perfectly (either because it does not match the experimental data or because from a theoretical point of view it lacks of "naturalness"). The search for $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$ has precisely as a final goal the confirmation or discarding of these models (NP and SM), and a summary on which particular aspects of the models both channels are more sensitive to is also given.

2.2.1 Motivations for new physics

2.2.1.1 Unexplained data

There are several measurements which produce tensions with the SM or are simply unexplainable with it, in the sense that it is hard to justify them using the information provided by the current accepted

theory. In this way, one of the logical requirements to the NP models is to give a better explanation to these measurements, when compared to the SM.

The most typically quoted example of unexplained measurement by the SM is the evidence of dark matter (DM), whose need to make sense of our Universe has been increasing for some time and which cannot be explained in the SM framework. In 1933 it was found that the velocity dispersion of galaxies in the Coma cluster was far too large to be supported by the luminous matter [26]. In the 1970s, the rotation curves of individual galaxies were measured, finding evidence of non-luminous matter [27, 28]. More recently, evidence for non-luminous matter has been supplemented by data from gravitational lensing, such as, for instance, that in the Bullet Cluster [29]. Candidates for cold DM are primordial black holes [30, 31], axions [32, 33, 34], and weak interacting massive particles (WIMPs), such as the lightest supersymmetric particle (LSP). A very complete review on DM can be found in [35].

Also, the neutrino flavour oscillation evidence [36] implies that this particle is massive, in contradiction with SM postulates, although this can be solved in principle by adding some mechanism like seesaw [37], without changing too much the basic ideas of the SM.

Another example of unexplained measurement is the prevalence of matter over anti-matter in our Universe, known as the Baryogenesis problem. In order to account for this prevalence, three conditions that the baryon-generating interaction should satisfy were proposed in 1967 [38]. Among these is a CP-violation rate larger than the one included in the SM. So it can be said that the SM is unable to explain the matter-antimatter unbalance.

Finally, the measurement of the anomalous magnetic dipole moment of the muon [39, 40] (which deviates more than 3σ from the SM prediction), the evidence for CP violation asymmetry in the charm sector [41] or the evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ [42] decays (again more than 3σ above the SM prediction) produce also some tensions, although to a lesser extent than the previous examples.

2.2.1.2 SM theoretical problems

Apart from the explained experimental motivations, there are a series of theoretical aspects that make the SM unsatisfactory. A summary of the most important is shown next [35, 43]:

- The gauge hierarchy problem: this can be understood simply as the reason that makes the physical Higgs boson mass m_h so small. If three of the most fundamental constants (the speed of light c, Planck constant h, and Newton gravitational constant G_N) are combined properly, the result has dimensions of mass, the resulting quantity known as the Planck mass: $M_{Pl} = \sqrt{\hbar c/G_N} = 1.2 \times 10^{19}$ GeV. Therefore, one would expect dimensionful parameters to be either 0, if enforced by a symmetry, or of the order of M_{Pl}. In the SM, electroweak symmetry is broken, and the Higgs boson mass is non-zero. The gauge hierarchy problem is the question of why $m_h \sim 100$ GeV $\ll M_{Pl}$.
- Charge quantisation: this refers to the experimental fact that the electron charge is equal but opposite in sign to the proton charge ($|Q_e + Q_p|/e < 1.0 \times 10^{-21}$ [5]), not explained in the SM.
- The strong CP problem: The SM Lagrangian includes the term $g_3^2\theta_3/(32\pi^2)\epsilon^{\mu\nu\rho\sigma}G^\alpha_{\mu\nu}G^\alpha_{\rho\sigma}$, where g_3 is the coupling of the strong interactions, θ_3 is an angle parameter, $\epsilon^{\mu\nu\rho\sigma}$ is the totally anti-symmetric 4-index tensor, and $G_{\mu\nu}$ is the gluon field strength. This term contributes to CP-violating flavour-conserving observables, but experimentally θ_3 is known to be extremely small ($\theta_3 \lesssim 10^{-13}$, extracted from [44]).
- Fermion masses and mixing angles: the fermion mass spectrum ranges from ∼170 GeV, for the case of the top-quark, to $\sim 10^{-3}$ GeV [5], for the case of the electron. This reflects in the SM Lagrangian in a large number of free parameters in the Yukawa part, which makes the SM look more like an effective low energy theory.
- Differences in the running of the couplings of the interactions. As shown in figure 2.8 (left), the gauge couplings of the electromagnetic, weak and strong interactions do not perfectly match at high energies, as it would seem to be *natural*.
- Gravity cannot be explained in the framework of SM (due to the lack of a quantum description for it).

2.2.2 New physics models

Several NP models will be briefly described now. The models have been grouped following [43, 1, 45].

2.2.2.1 Most general extensions to SM

2.2.2.1.1 Minimal Flavour Violation The definition of MFV (Minimal Flavour Violation) [46, 47] gathers models which require that the dynamics of flavour and CP violation are governed by the known structure of the Yukawa couplings.

The MFV condition can be constrained further (CMFV) [48, 49, 50] by also imposing that the only relevant operators in the effective Hamiltonian below the weak scale are those that are also relevant in the SM.

2.2.2.1.2 Two Higgs Doublet Models The Two Higgs Doublet Model (2HDM) [51, 52] is an extension of the SM containing a larger Higgs sector, coming from two Higgs doublets both with a VEV$\neq$0. In an analogous way to what it was seen in equation 2.4, the VEV of both doublets can be written as:

$$\langle\phi_a\rangle_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu_a \end{pmatrix} \qquad \langle\phi_b\rangle_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu_b \end{pmatrix} \tag{2.24}$$

The ratio $\tan\beta = \frac{\nu_a}{\nu_b}$ will be a very important free parameter of the theory (both VEV are not independent, since $\nu_a^2 + \nu_b^2 = \nu_{SM}^2 \sim (174\text{ GeV})^2$).

The spontaneous symmetry breaking in the case of 2HDM yields 5 physical Higgs bosons (two charged and three neutral), in comparison with SM, where only 1 physical Higgs boson is present.

2.2.2.1.3 Standard Model with fourth generation One of the simplest extensions of the SM is the addition of a sequential fourth generation of quarks and leptons (SM4) [53]. This model has potential to solve some of the current experimental puzzles, while it goes over the measurement at LEP that $N_\nu = 3.00 \pm 0.08$ [5] by assuming 4th generation neutrinos with masses above $M_{Z^0}/2$ [54] (the measurement at LEP was obtained from the total Z^0 decay width).

2.2.2.2 Supersymmetry

The symmetry which requires the theory to be invariant under the transformation of fermions to bosons (and vice versa) is called SuperSymmetry (SUSY). This implies that the model should have the same number of fermions than bosons, establishing for each SM fermion the corresponding "sfermion" (*selectron*, *squark*...) and for each boson the corresponding fermion, named by adding the suffix ino (*gluino*, *photino*, *neutralino*...).

Since no SUSY particles have been observed so far in the colliders, if this symmetry exists it has to be broken at low energies (compared to the Planck scale). The way this SUSY-breaking occurs can be very different, and different submodels arise depending on this. Within supersymmetry, it is also important to mention the *R-parity*, which is defined so that the SM particles have $R = 1$ and their superpartners $R = -1$.

SUSY has the capability to solve some of the SM problems seen in section 2.2.1. In this way, if R-parity invariance is imposed, the Lightest Supersymmetrical Parity (LSP) immediately becomes a clear DM candidate. SUSY also solves the problem in the running of the couplings of the interactions. As shown in figure 2.8 (right), within the SUSY framework the couplings nicely match at high energies.

A wide review on supersymmetry can be found in [55]. Two particular realisations of SUSY will be quickly described now.

2.2.2.2.1 Minimal Supersymmetric Standard Model The MSSM (Minimal Supersymmetric Standard Model) is one of the most popular SUSY models and arises simply by imposing supersymmetry on the SM [56].

The Higgs structure in MSSM is similar to 2HDM, so that $\tan\beta$ is also a very important parameter for MSSM. Examples of different variations of MSSM are:

- SF (SUSY flavour models), which contain flavour symmetries that allow for some understanding of the flavour structures in the Yukawa couplings and in SUSY soft-breaking terms, adequately suppressing FCNC and CP violating phenomena and solving SUSY flavour and CP problems [57].
- mSUGRA (minimal Super Gravity), which combines the principles of supersymmetry and general relativity, so that SUSY breaking is mediated by gravity [58].

- CMSSM (Constrained MSSM), in which the boundaries applied are similar to those of mSUGRA, with mSUGRA being a bit more restrictive [59].
- NUHM (Non Universal Higgs Masses), in which the SUSY breaking is also mediated by gravity, but the universality of scalar masses is relaxed for the Higgs doublets [60].
- RPV MSSM (R Parity Violating MSSM), similar to MSSM, but allowing the R-parity to be violated [61]. These models have the disadvantage of not predicting a DM candidate, since the LSP can also decay to ordinary matter.

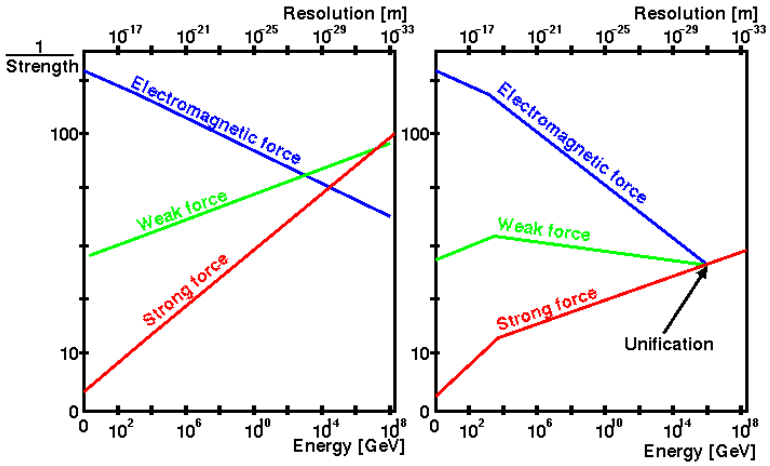

Figure 2.8: Evolution of the three gauge couplings as a function of the energy in the SM (left) and SUSY (right). The figure has been taken from [62].

2.2.2.2.2 Next to Minimal Supersymmetric Standard Model The NMSSM (Next to Minimal Supersymmetric Standard Model) adds to MSSM an additional gauge singlet in order to solve some of the theoretical problems of MSSM [63]. The particle content of NMSSM adds two additional Higgs bosons to MSSM, while in the fermionic part one additional neutralino, called singlino, is also added. An extensive review on NMSSM can be found in [64].

NMSSM has as well different variations, among which is the Constrained NMSSM (CNMSSM) [65], which inputs soft SUSY-breaking parameters that are universal at a high scale. With this, the number of unknown parameters of NMSSM is reduced to a handful.

2.2.2.3 Higgless and composite Higgs

2.2.2.3.1 Littlest Higgs Models Little Higgs models (LH) [66] attempt to solve the hierarchy problem by using the idea of light composite Higgs and without using SUSY. In these models the Higgs arises as a pseudo-Goldstone boson produced by the breakdown of some global approximate symmetry at the TeV scale.

In order to make this model consistent with electroweak precision tests and simultaneously having its new particles in the reach of the LHC, a discrete symmetry, "T-parity", has been introduced [67, 68], in which is called LHT (Littlest Higgs with T parity). T-parity also provides DM candidates, so that under T-parity all SM particles are even.

2.2.2.3.2 Technicolor Technicolor (TC) models remove the need of a Higgs boson by using composite scalar fields as mechanism for EWSB. Moreover, in order to produce quark and lepton masses, Technicolor has to be "extended" by additional gauge interactions (ETC). An extensive review on Technicolor models can be found in [69].

The simplest TC models suffer from inconsistency with experimental observations, so a way of allowing Technicolor to be consistent with observed phenomenology is through the introduction of a $t - \bar{t}$ condensate due to a new strong interaction in the top system [70]. The combination of this interaction with ETC is called Topcolor assisted Technicolor (TC2).

2.2.2.4 Other models

2.2.2.4.1 Extra-dimensions The possibility of adding extra dimensions (EDs) to the usual 4-D space-time was first proposed by Kaluza [71] and Klein [72] in order to unify electromagnetism and gravity into a common origin. In the Kaluza-Klein picture, the extra dimensions are compact, with a microscopic size L ensuring that space-time becomes four dimensional at distances $\ll L$. However, in the so called braneworld picture, the mechanism for hiding extra dimensions is performed by trapping the ordinary matter into a 3D submanifold (brane) embedded in the fundamental multidimensional space so that the extra dimensions can be large or even infinite. Depending on the structure of EDs, the models have different properties and features. One of the most popular among EDs models are those with warped EDs or Randall-Sundrum (R-S) models [73]. These provide a geometrical explanation of the hierarchy problem and can also naturally generate the hierarchies in the fermion masses and mixing angles [74].

2.2.2.4.2 Axions One of the most popular solutions to the strong CP problem, introduced in section 2.2.1, was proposed by Peccei and Quinn [32]. Essentially it consists in the promotion of the parameter θ of the SM Lagrangian to a field $a(x)$, the axion field, assumed to be a Goldstone boson arising from the spontaneous breaking of a $U(1)$ symmetry, the PQ symmetry. The Peccei-Quinn mechanism has as a testable prediction the existence of a new particle, the axion, whose mass can be calculated to be in the range 100 keV to 10^{12} eV. As said before, the axion is one of the candidates to explain DM. Finding this particle is the goal of several experiments currently taking place.

2.2.3 $B_s^0 \to \mu^+\mu^-$ in New Physics models

$B_{(s)}^0 \to \mu^+\mu^-$ decays are among the "golden modes" for searches of physics beyond the SM [75], due to their very high sensitivity to several NP models, particularly in the scalar sector. Indeed, the measurement of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ would severely constraint their phase space or even completely discard them, depending on its actual values.

A complete review on the sensitivity of NP models to the $\mathcal{B}(B_{(s)}^0 \to \mu^+\mu^-)$ can be found in [1, 45]. Some of the most important ideas on the NP constraints from these channels, particularly for $B_s^0 \to \mu^+\mu^-$, are summarised next:

- Not only the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are sensitive to NP, but also its ratio. For instance, in CMFV models the ratio $\frac{\mathcal{B}(B^0\to\mu^+\mu^-)}{\mathcal{B}(B_s^0\to\mu^+\mu^-)}\big|_{CMFV}$ is expected to be $\sim$ equal to $\frac{\mathcal{B}(B^0\to\mu^+\mu^-)}{\mathcal{B}(B_s^0\to\mu^+\mu^-)}\big|_{SM}$.
- The $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in many of the models in which two Higgs doublets are present (such as 2HDM or MSSM) is very sensitive to $\tan\beta$ (defined in section 2.2.2.1.2). In particular, for 2HDM, $\mathcal{B}(B_s^0 \to \mu^+\mu^-) \propto (\tan\beta)^4$, and in MSSM $(\tan\beta)^6$ contributions are obtained.
- Constrained fits to SUSY models with few free parameters, such as CMSSM and NUHM1[5], are predictive enough to provide expectations for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ [76]:

$$\frac{\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{CMSSM}}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{SM}} \approx 1.2^{+0.8}_{-0.2}$$
$$\frac{\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{NUHM1}}{\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{SM}} \approx 1.9^{+1.0}_{-0.9}$$

- Several new particles can mediate these decays, even at tree level. Examples of $B_s^0 \to \mu^+\mu^-$ Feynman diagrams in different NP model are shown in figure 2.9.

Table 2.2 advances different NP scenarios depending on the observed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. For $B^0 \to \mu^+\mu^-$, more information can be found in [45].

2.2.4 $K_S^0 \to \mu^+\mu^-$ in New Physics models

The SD contribution to the $\mathcal{B}(K_S^0 \to \mu^+\mu^-)$ was shown in equation 2.22 to be proportional to $\bar{\eta}$. Within SM, this contribution was seen to be at the level of 10^{-13}. However, the present constraints on $\bar{\eta}$ derived

[5] NUHM1 is particular realisation of NUHM.

Scenario	Would point to
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \gg$SM	Big enhancements from NP in scalar sector, SUSY with high $\tan\beta$
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \neq$SM	SUSY, EDs, LHT, TC2
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \sim$SM	rule out regions of parameter space that predict sizeable departures from SM
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \ll$SM	NP in scalar sector, but MSSM ruled out or quite unlikely. NMSSM good candidate

Table 2.2: Implications for NP of different measured $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$

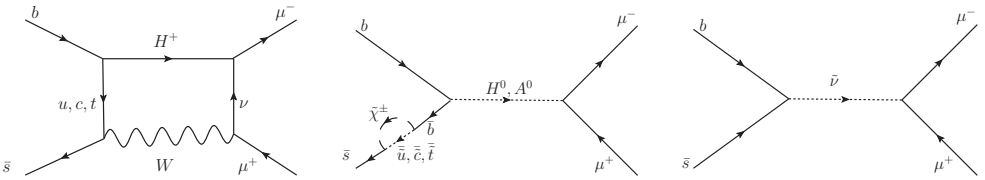

Figure 2.9: Diagrams contributing to $B_s^0 \to \mu^+\mu^-$ in NP. Left, in 2HDM models. Center, MSSM. Right, models with R-parity violation. As explained, in RPV $B_s^0 \to \mu^+\mu^-$ can arise at tree level.

only from $\Delta S = 1$ FCNC processes are rather weak. This implies that new-physics scenarios where $\mathcal{B}(K_S^0 \to \mu^+\mu^-)\,|_{SD}$ is at the 10^{-11} level are perfectly allowed and also that enhancements above 10^{-10} are less likely [22].

Figure 2.10 shows the current bounds in the $\bar{\eta}$ vs. $\bar{\rho}$ plane (where both these parameters are related to the CKM matrix elements trough the Wolfenstein parameterisation). It is interesting to see that the bounds from the $\mathcal{B}(K_L^0 \to \mu^+\mu^-)$ (which is proportional to $\bar{\rho}$) are orthogonal to those of $\mathcal{B}(K_S^0 \to \mu^+\mu^-)$ (which as just said is proportional to $\bar{\eta}$). It is also important to remark that, contrary to the $K_L^0 \to \mu^+\mu^-$ case, the dominant dispersive long-distance in $K_S^0 \to \mu^+\mu^-$ is unambiguously determined at the lowest order in the chiral expansion, as seen in section 2.1.4.1. With this, any significant excess in the $\mathcal{B}(K_S^0 \to \mu^+\mu^-)$ with respect to the SM expectation could only be explained by NP effects. Figure 2.10 also shows that a limit in the $\mathcal{B}(K_S^0 \to \mu^+\mu^-)$ could also constraint the phase space of other important models of the kaon sector, such as $K^+ \to \pi^+\nu\bar{\nu}$.

Finally, [22] states that bounds on $\mathcal{B}(K_S^0 \to \mu^+\mu^-)$ close to 10^{-11} could be translated into interesting model-independent bounds on the CP-violating phase of the $s \to dl^+l^-$ amplitude. These bounds could be very useful to discriminate among NP scenarios if other modes, such as $K^+ \to \pi^+\nu\bar{\nu}$, indicated a non-standard enhancement of the $s \to dl^+l^-$ transition. In [21] $K_S^0 \to \mu^+\mu^-$ has also been suggested as a possible way to look for new light scalars.

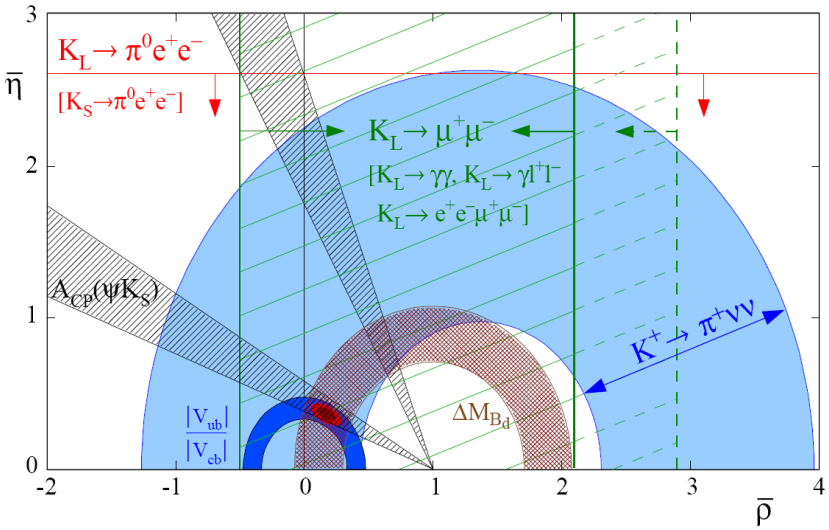

Figure 2.10: Current bounds in the $\bar{\eta}$ vs. $\bar{\rho}$ plane. The bounds in $\mathcal{B}(K_S^0 \to \mu^+\mu^-)$, which is proportional to $\bar{\eta}$, and in $\mathcal{B}(K_L^0 \to \mu^+\mu^-)$, which is proportional to $\bar{\rho}$, are orthogonal. Figure taken from [22].

3

The LHCb experiment at the LHC

3.1 The LHC

The Large Hadron Collider (LHC) is currently the man-made world most powerful particle accelerator located in the world most important particle physics laboratory, CERN [77]. Being situated in a 27 km ring buried deep below the countryside on the outskirts of Geneva (Switzerland), the LHC is installed in the tunnel that previously housed the Large Electron Positron collider (LEP).

The aim of the LHC project is to test the Standard Model and, even more, to look for signals of physics beyond it. In the LHC, protons currently collide at a centre-of-mass energy of $\sqrt{s} = 8$ TeV, the highest ever achieved in *pp* collisions in man-made colliders. During 2011, when the data used for this thesis was taken, the collisions took place at $\sqrt{s} = 7$ TeV. The design energy for the collisions (expected at the end of 2014 [78]) is $\sqrt{s} = 14$ TeV [79, 80]. At these energies, new discoveries such as, e.g., the existence of SUSY, are being tried to be made. Also, the only unobserved particle in the Standard Model, the Higgs-boson, is expected to be produced and detected (with first hints of it in 2011 [81, 82] and a likely observation in 2012 [83, 84]). The LHC provides a perfect facility to search for these new particles. Other research topics concern high-precision B physics, and the study of a new phase of matter produced in heavy-ion collisions. For the latter, the LHC is filled in dedicated runs with heavy ions (e.g., Pb) instead of protons.

During 2011, the proton beams in the circular ring were accelerated in opposite directions to an energy of 3.5 TeV. Before reaching that energy, they had passed through a chain of preaccelerators. At the last stage of the preacceleration, the Super Proton Synchrotron (SPS) delivered 450 GeV protons to the LHC, where they were boosted to the final energy. Figure 3.1 shows the CERN acceleration complex, including the LHC and all the preaccelerators. A real photo of the accelerator installed at the ring can be seen at figure 3.2.

At the LHC running energies, protons require a very intense magnetic field to maintain their orbit (8.33 T at the nominal 7 TeV [79, 80]). This field is provided by superconducting magnets. As the two proton beams travel in opposite directions through the ring, separate beam pipes with opposite magnetic field directions are needed. A two-in-one solution is chosen, where the magnet coils surrounding the two beam channels are firmly embodied inside the same iron yoke. The whole superconducting magnet is placed inside a cryostat, containing superfluid helium with a temperature of 1.9 K. Figure 3.3 shows the magnetic flux in the LHC yoke.

The four main experiments at the LHC are located at each of the four interaction points. Here, the beams cross over to the other beam pipe and collide under a small angle. Figure 3.1 shows the location of the four experiments along the LHC ring. ATLAS [85] and CMS [86] are two general-purpose experiments. Both are central detectors constructed by large, international collaborations. Their main physics goals are the search for the Higgs boson and SUSY particles. In addition, these experiments plan to study B physics, heavy-ion collisions, and many other interesting phenomena. The ALICE [87] experiment focuses on studying strongly interacting matter at the extreme energy densities in heavy-

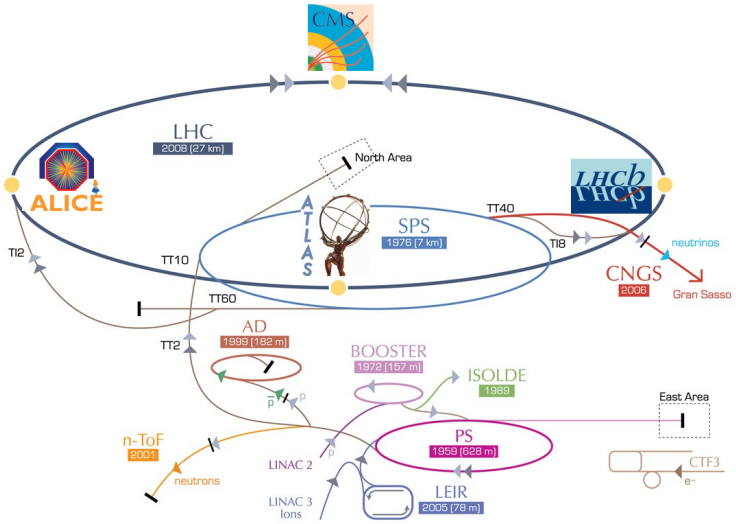

Figure 3.1: Display of the CERN accelerator complex, including the LHC and its four experiments

Figure 3.2: The LHC accelerator installed at the 27 km ring, 100 m underground.

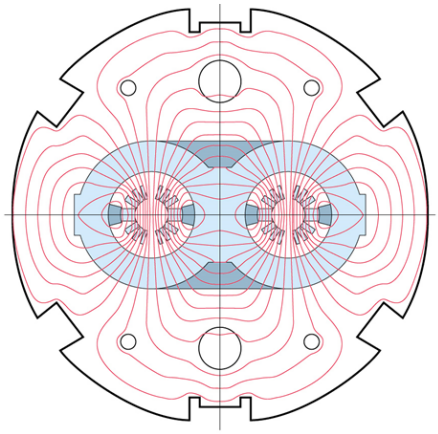

Figure 3.3: Magnetic flux in the LHC yoke.

ion collisions, and performing measurements of the phase transition between hadronic matter and the quark-gluon plasma. The ALICE detector is designed to cope with the higher particle multiplicities that are characteristic for heavy-ion collisions. Finally, LHCb [88] thought for rare decays and CP violation measures, will be treated in more detail in section 3.2. Figure 3.4 shows pictures of the four LHC experiments.

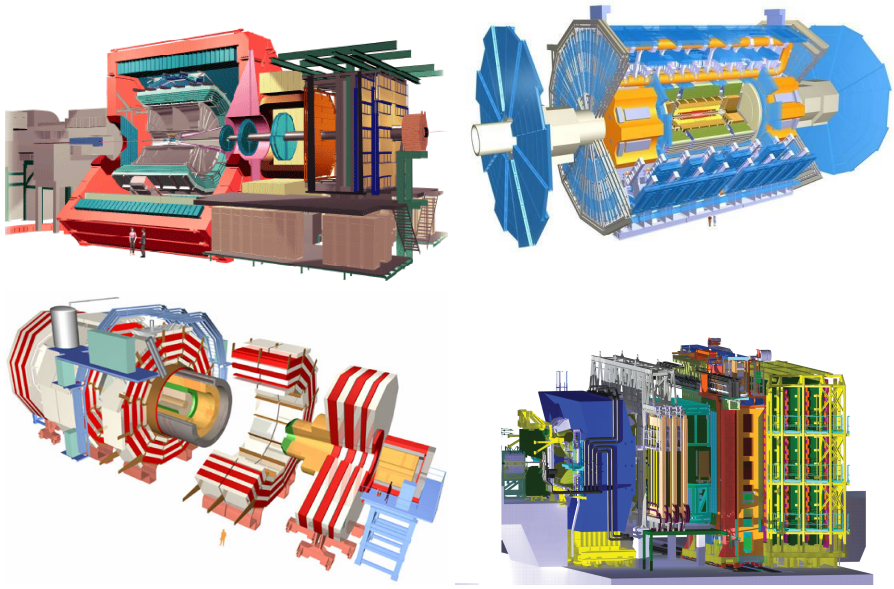

Figure 3.4: LHC experiments (not to the same scale). Top left, ALICE; top right, ATLAS; bottom left, CMS; bottom right, LHCb.

3.2 LHCb

LHCb is one of the four big experiments taking place at the LHC. It is currently composed by more than 700 scientists from 52 universities and laboratories from 15 countries. As ALICE, and in opposition to ATLAS and CMS, LHCb is an specific purpose experiment: study of CP violation and rare decays. Examples of the key physics measurements by LHCb can be found in [89]. A review on latest LHCb rare decays results will be given in chapter 4.

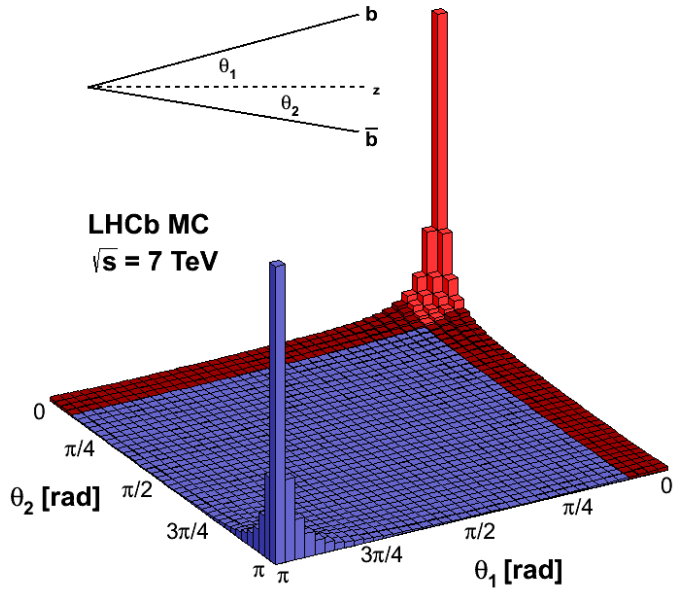

Figure 3.5: Angular distribution of b and $\bar{b}$ quarks in collisions at $\sqrt{s} = 7$ TeV.

LHCb experiment is thought to work at a instant luminosity of $2 - 5 \cdot 10^{32}$ cm^{-2}·s^{-1}, smaller than LHC nominal 10^{34} cm^{-2}· s^{-1}. This is achieved using larger β^{*1} values than for other LHC detectors, therefore less focusing of the beams. The reason to do so is easing the correct identification of the point where the proton–proton collision took place (primary vertex or PV) and the point where other short-lived but flying particles decayed (secondary vertex or SV), as this is essential for the physics in the experiment.

3.2.1 Detector

LHCb is a single-arm spectrometer with a forward angular coverage from approximately 10 mrad to 300 (250) mrad in the bending (non-bending) plane. The choice of the detector geometry is justified by the fact that at high energies both the b and $\bar{b}$ hadrons are predominantly produced in the same forward or backward cone (see figure 3.5).

The layout of the LHCb spectrometer is shown in figure 3.6. The right-handed coordinate system adopted has the z axis along the beam, and the y axis almost along the vertical (since there exists a ∼3.7 mrad tilt compared to the actual geometrical vertical).

Intersection Point 8 of the LHC, previously used by the DELPHI experiment during the LEP time, has been allocated to the LHCb detector. A modification to the LHC optics, displacing the interaction point by 11.25 m from the centre, has permitted maximum use to be made of the existing cavern for the LHCb detector [90] components. Main elements of LHCb are:

- a spectrometer magnet, a warm dipole magnet providing an integrated field of 4 T·m, described in section 3.2.1.1;
- a vertex locator (VELO) system (including a pile-up veto counter), explained in section 3.2.1.2.1;
- a Tracking System, made of the Tracker Turicensis (a silicon microstrip detector, TT) in front of the spectrometer magnet, and three tracking stations downstream the magnet, made of silicon microstrips in the inner parts (Inner Tracker or IT) and of gas straws for the outer parts (Outer Tracker or OT), to be seen in section 3.2.1.2.2;

[1] amplitude modulation of the beam in the interaction point, directly related to the size of the beam.

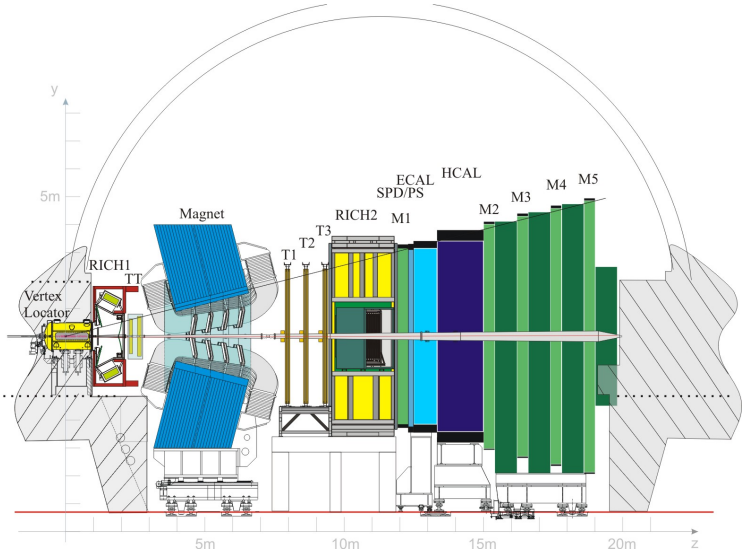

Figure 3.6: Side view of LHCb

- two Ring Imaging Cherenkov counters (RICH1 and RICH2) using Aerogel, C_4F_{10} and CF_4 as radiators, to achieve excellent p–K separation in the momentum range from 2 to 100 GeV/c, described in section 3.2.1.3.1;
- a Calorimeter system composed of a Scintillator Pad Detector and Preshower (SPD/PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL), explained in section 3.2.1.3.2;
- a Muon detection System composed of MWPC (except in the highest rate region, where triple-GEMs are used), to be seen in section 3.2.1.3.3.

Apart from the magnet, the just described detectors can be classified according to its function. In this way, the VELO and tracking system are generally considered "Tracking" detectors, while RICH, Calorimeters and Muon System "PID" (Particle Identification) detectors.

3.2.1.1 Magnet

The LHCb dipole magnet [91] provides a magnetic field of 4 T·m that bends the charged particles in the horizontal plane of the detector with the idea of allowing the measurement of their momenta. The measurement covers the forward acceptance of ±250 mrad vertically and of ±300 mrad horizontally.

The dipole is composed by a Fe yoke surrounded by two identical coils of conical saddle shape produced of pure Al-99.7. Figure 3.7 (left) shows a picture of the dipole installed at the LHCb experimental site.

The magnetic field provided by the dipole must be known with excellent precision, in order to yield a momentum resolution as good as possible. The precision of the measurement obtained for the field mapping in the tracking volume is about 4×10^{-4}. Figure 3.7 (right) shows the B_y component (which is the most relevant) as a function of the z coordinate. It is also important to say that for the measurement of CP asymmetries and in order to control the systematic effects of the detector, the direction of the magnetic field is changed periodically.

3.2.1.2 Tracking detectors

3.2.1.2.1 VELO The Vertex Locator (VELO) [90, 92, 93] contains 21 stations, positioned along and perpendicular to the beam axis. Figure 3.8 (left) shows a partial view of the VELO as seen from above. Two types of silicon sensors are used: one measures the r coordinate with circular strips centered around the beam axis, the other measures the ϕ coordinate with straight, almost radial strips (including a stereo-angle built in). The half-disc sensors, shown in figure 3.8 right, are arranged in pairs of r and ϕ sensors and mounted back-to-back. The radius of each module is around 42 mm. The minimum

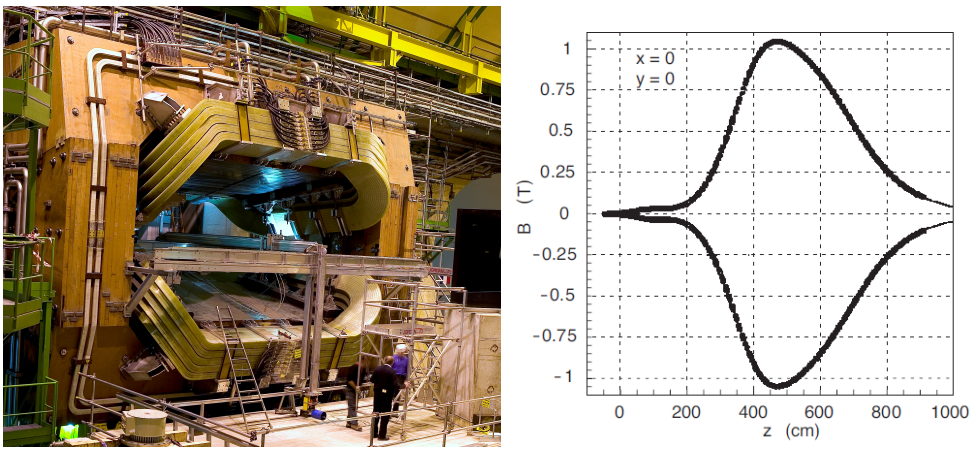

Figure 3.7: LHCb magnet dipole. Left figure shows a picture of the detector installed at the LHCb experimental site. Right figure shows the B_y component of the magnetic field as a function of the z coordinate

strip pitch in the sensors (at innermost radius) is 38 μm, with a linear increase up to 101.6 μm. With this, a spatial cluster resolution of about 4 μm is achieved for 100 mrad tracks (corresponding to the smallest strip pitch region). The acceptance of VELO covers the pseudorapidity range $1.6 < \eta < 4.9$ for particles coming from primary vertices in the range $-10.6 < z < 10.6$ cm.

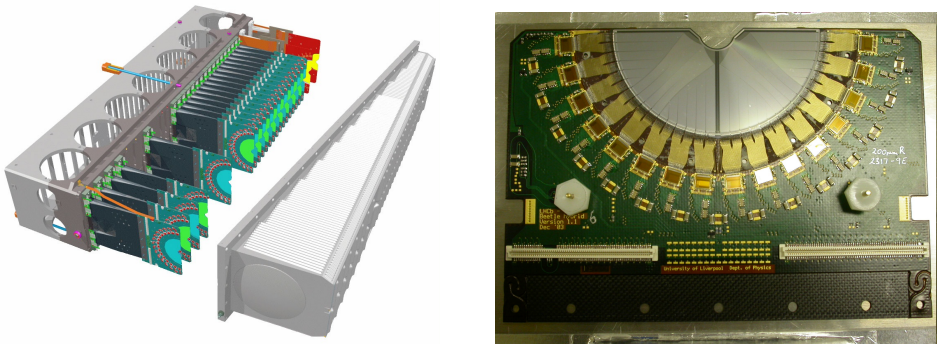

Figure 3.8: LHCb Vertex Locator. Left, overview of VELO. Right, photograph of one of the modules.

The $r-\phi$ geometry has the advantage that it directly gives a projection in the rz plane by using only r measurements. In this projection, forward-going tracks with a high impact parameter with respect to the production vertex are easily identified.

Moreover, two dedicated silicon stations containing r sensors only are located upstream of the VELO and were initially conceived for pile-up veto in the trigger, although this idea has not been finally implemented.

The sensitive area of the sensors starts at 8 mm from the beam axis, such that the first measurement of the track is as close to the primary vertex as possible. The shorter the extrapolation of a track from its first measurement to the interaction region, the smaller is the error on the reconstructed position of the vertex. This proximity requirement implies that the sensors must be retractable during beam injection. In order to avoid severe radiation damage or even the beam to go straight through the sensors, a minimal distance of 3 cm is required when LHC is being filled. Consequently, the VELO is designed so that the two detector halves can be moved away from the beam in the horizontal direction.

3.2.1.2.2 Tracking system As explained, the LHCb tracking system is composed by the TT [93], right upstream the magnet, and by three tracking stations downstream the magnet and before

RICH2. The tracking stations have two different substructures: the IT [94] in the most inner part and the OT [95] surrounding it. Since the technology of the TT and the three IT modules is the same (silicon microstrips), both are typically grouped in what is called the Silicon Tracker (ST).

The four ST stations use silicon microstrip sensors with a strip pitch of about 200 μm (giving a single hit resolution of 50 μm). In order to get a 3D reconstruction, each of the stations includes four detection layers in an (x-u-v-x) arrangement with vertical strips in the first and the last layer and strips rotated by a stereo angle of -5° and +5° in the second and the third layer. As for the acceptance, TT is 150 cm wide and 130 cm high and covers the full acceptance of the experiment. The IT covers a 120 cm wide and 40 cm high cross shaped region in the centre of the three tracking stations. Figure 3.9 shows pictures of both the IT and TT.

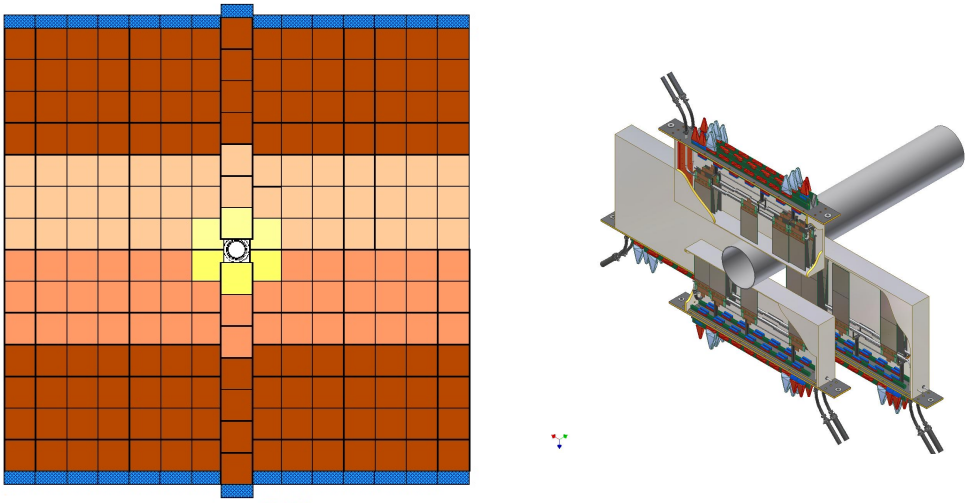

Figure 3.9: Silicon tracker of LHCb. Left, layout of one of the TT layers. Right, one of the three IT modules.

The OT is a drift-time detector using as counting gas a mixture of Argon (70%) and CO_2 (30%), with which a drift time below 50 ns and a drift-coordinate resolution of 200 μm is achieved. The detector modules are arranged in three stations and, similarly to the ST, each station consists of four layers, so that the modules in first and last layers are oriented vertically, while those in the center are tilted by ±5° with respect to the vertical. The total active area of a station is 6×5 m. Figure 3.10 shows a photograph of the three OT modules at the LHCb site.

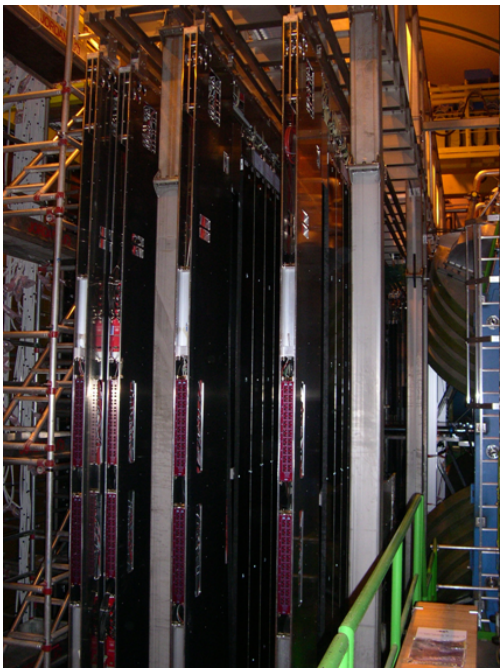

Figure 3.10: Real image of the three OT modules installed at the LHCb experimental site.

3.2.1.3 PID subdetectors

3.2.1.3.1 RICH LHCb includes two Ring Imaging Cherenkov detectors [96], aiming at different momentum ranges. The upstream detector, RICH1, covers the low momentum charged particle range ∼1- 60 GeV/c using aerogel and C_4F_{10} radiators, while the downstream detector, RICH2, covers the high momentum range from ∼15 GeV/c up to and beyond 100 GeV/c using a CF_4 radiator. RICH1 has a wide acceptance covering the full LHCb acceptance from ±25 mrad to ±300 mrad (horizontal) and ±250 mrad (vertical). RICH2 is located downstream of the magnet and has a more limited angular acceptance of ±15 mrad to ±120 mrad (horizontal) and ±100 mrad (vertical). Figure 3.11 shows a picture of both RICH1 and RICH2.

In both RICH detectors the focusing of the Cherenkov light is accomplished using a combination of spherical and flat mirrors to reflect the image out of the spectrometer acceptance. Hybrid Photon Detectors (HPDs) are used to detect the Cherenkov photons in the wavelength range 200-600 nm. The HPDs are surrounded by external iron shields. Figure 3.12 represents a typical event at RICH, both with the event display and the pattern of rings produced.

3.2.1.3.2 Calorimeter system The LHCb calorimeter system [97] has the function of selecting transverse energy hadron, electron and photon candidates for the first trigger level (L0) and providing the identification of electrons, photons and hadrons as well as the measurement of their energies and positions.

The calorimeter system is composed by a Preshower detector (PS) and a Scintillator Pad Detector (SPD) plane before the PS, and an Electromagnetic Calorimeter (ECAL) to which it follows a Hadron Calorimeter (HCAL). The purpose of having these four substructures is properly discriminating between hadrons, electrons/positrons and photons, since the energy deposition in each of the calorimeter components will depend on the nature of the particles. Figure 3.13 left shows an sketch of the energy loss by different particles in SPD, PS, ECAL and HCAL. The Calorimeter PID (which will be explained in section 3.2.3.3) makes use of this information to correctly identify the particles.

In the whole calorimeter system, the scintillation light is transmitted to Photo– Multipliers (PMTs), that turn this light into an electric signal. As far as the subcomponents are concerned, the SPD/PS detector consists of a 15 mm lead converter, that is sandwiched between two almost identical planes of rectangular scintillator pads. The sensitive area of the detector is 7.6 m wide and 6.2 m high. The ECAL employs alternating scintillating tiles and lead plates, with overall dimensions 7.8 × 6.3 m. Finally, the

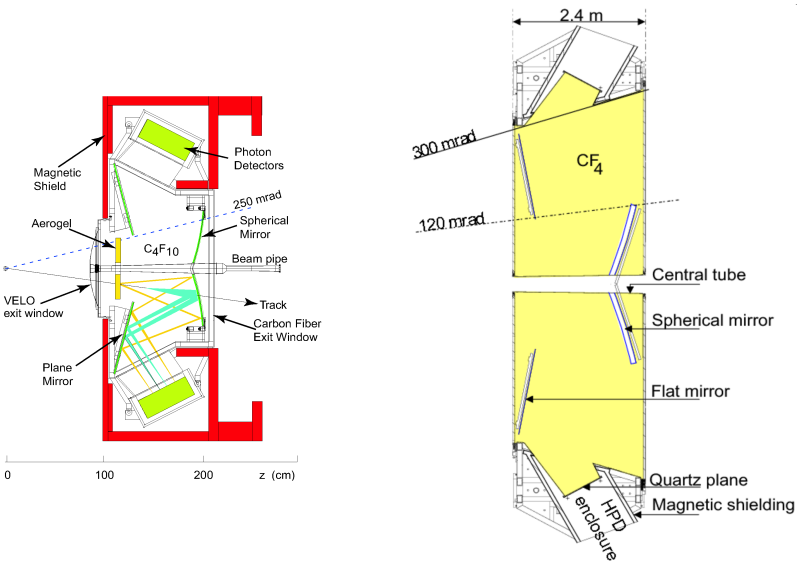

Figure 3.11: Layout of the RICH subdetectors. Left, RICH1. Right, RICH2.

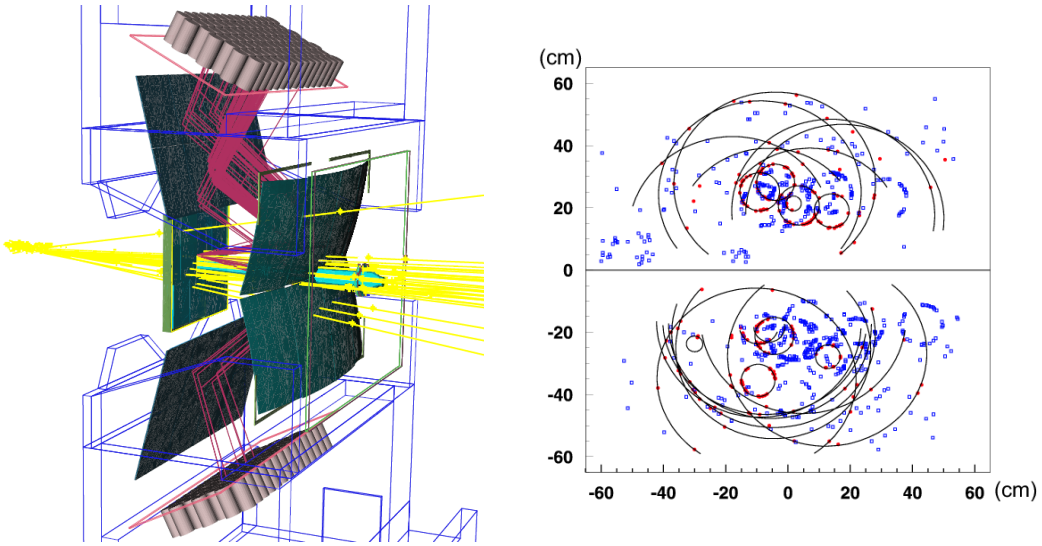

Figure 3.12: Example of typical event at RICH1. Left, event display with particles going through detector and Cherenkov radiation produced. Right, pattern of rings produced by the different particles.

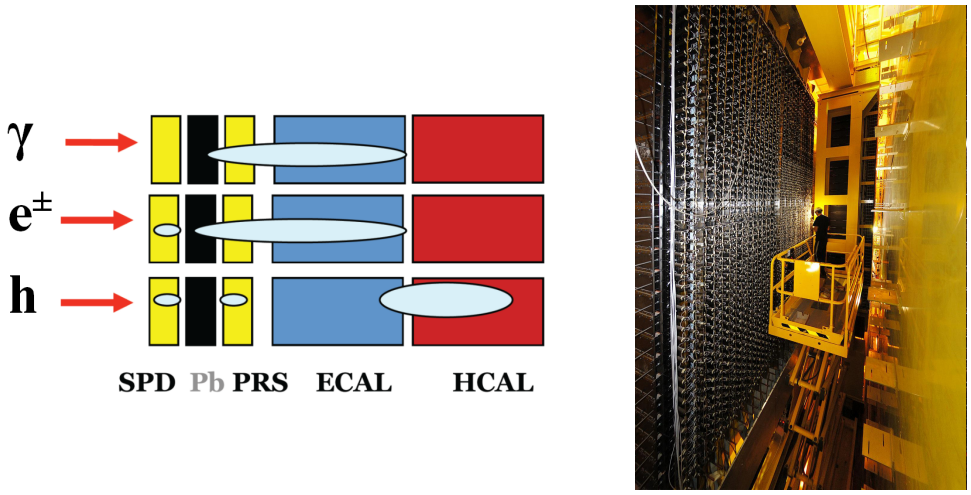

Figure 3.13: Calorimeter system of LHCb. Left, energy deposition on the subelements of the calorimeter system depending on the nature of the particle (h referring to hadrons). Right, view of the ECAL PMTs.

HCAL, consists of thin iron plates interspaced with scintillating tiles arranged parallel to the beam pipe. Its dimensions are 8.4×6.8 m. Figure 3.13 right shows a photograph of the PMTs of ECAL at the detector.

3.2.1.3.3 Muon detector The muon system for the LHCb experiment [98] consists of five tracking stations placed along the beam axis. The first station (M1) is placed in front of the calorimeter preshower, while the remaining four stations (M2, M3, M4 and M5) are located downstream the calorimeter, interleaved with three iron filters. Figure 3.14 left shows a photograph of the M5 station installed at the LHCb site.

The inner and outer angular acceptances of the muon system are 20 (16) mrad and 306 (258) mrad in the bending (non-bending) plane, similar to that of the tracking system. This provides a geometrical acceptance of about $\sim 20\,\%$ for muons from b decays relative to the full solid angle. Each station is subdivided in four regions with dimensions and logical pad size which scales a factor of two from one region to the next (see figure 3.14 right).

Since the multiple scattering in the absorber increases with the distance from the beam axis, limiting the spatial resolution of the detector, the granularity of the detector varies accordingly. The logical pad dimension has been chosen such that its contribution to the p_T resolution is approximately equal to the multiple-scattering contribution.

The muon stations are equipped with Multi Wire Proportional Chambers [98] (MWPCs) operating with an Ar:CO_2:CF_4 (40%:55%:5 % in volume) gas mixture. The only exception to the MWPCs is the innermost region R1 of the station M1, in which the high rate of particles requires the use of a triple-GEM detector, which consists of three gas electron multiplier (GEM), using Ar:CO_2:CF_4 (45%:15%:40% in volume) as a gas mixture.

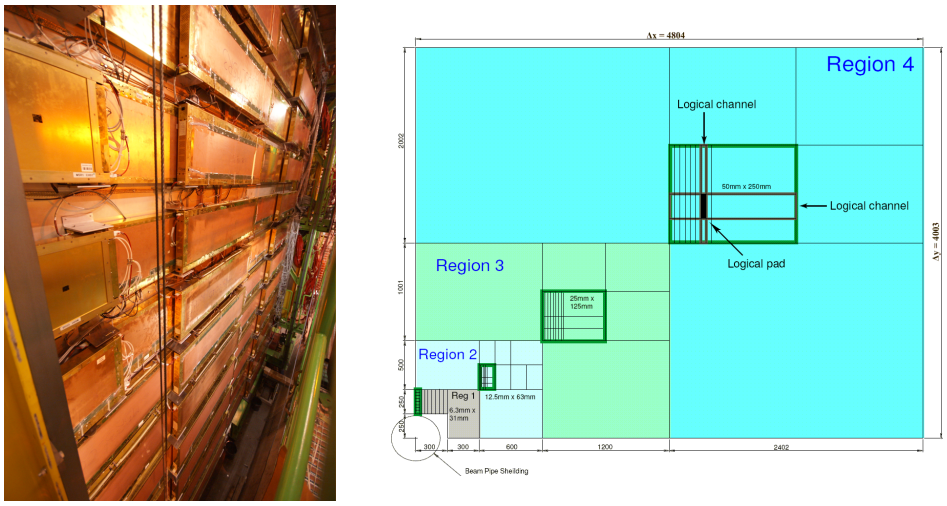

Figure 3.14: LHCb Muon System. Left, photo with the back view of M5. Right, front view of one quadrant of muon station 2, showing the dimensions of the regions. A sector is shown inside each region, defined by the size of the horizontal and vertical strips.

3.2.2 2011 experimental conditions

The analyses of this thesis are based on the data taken by LHCb from the LHC proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 7$ TeV between March and October 2011. This dataset corresponds to ~ 1.02 fb^{-1} of integrated luminosity.

The LHC machine started the operations in March by reaching very soon a peak luminosity L $\sim 1.6 \times 10^{32}$ cm^{-2} s^{-1} with 228 bunches (180 bunches colliding in LHCb) and an average number of pp visible interactions per crossing of $\mu \sim 2.5$. After the first 10 pb^{-1} collected by LHCb, the machine moved to the 50 ns bunch scheme and kept increasing the number of bunches by 144 every three fills, by reaching 1380 circulating bunches (1296 colliding bunches in LHCb) beginning of July. Since then the peak luminosity in LHCb was continuously leveled by changing the separation of the beams in order not to exceed $3 - 3.5 \times 10^{32}$ cm^{-2} s^{-1} corresponding to an average $< \mu > \sim 1.5$. Figure 3.15 shows the instantaneous luminosity at ATLAS, CMS and LHCb during a typical LHC run. As just said, the luminosity leveling yields a constant luminosity for LHCb.

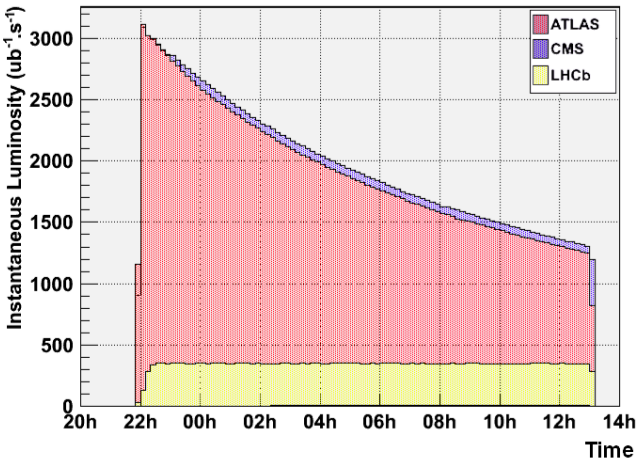

Figure 3.15: Instantaneous luminosity at ATLAS, CMS and LHCb vs. time during a typical LHC run. The luminosity leveling yields a constant luminosity for LHCb.

The considered data set was taken with a magnetic field both with positive ($B_y > 0$) (~ 446 pb^{-1}) and negative ($B_y < 0$) polarity (~ 595 pb^{-1}).

Figure 3.16 shows the integrated luminosity and the average value of proton-proton interactions per crossing (μ) as a function of the fill number for the period of data taking considered in this analysis.

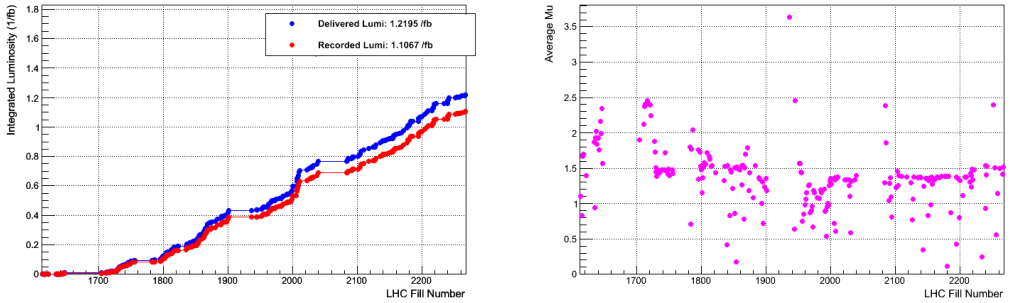

Figure 3.16: Integrated luminosity (left) and average number of *pp*-interactions per crossing (right) as a function of the fill number during 2011.

During the 2011 run, the luminosity increased very quickly to the regime value of $L = 3 - 3.5 \times 10^{32}$ cm^{-2} s^{-1} and kept constant by continuous levelling (see figure 3.17): about 1 fb^{-1} of data were collected in these conditions.

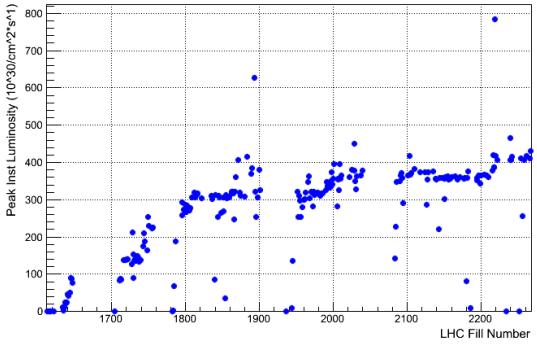

Figure 3.17: Peak luminosity as a function of the fill number during 2011.

3.2.3 LHCb analysis workflow

Data taken by LHCb from LHC proton–proton collisions, at a rate of several million events per second, need to be selected, prepared and distributed in the most possible efficient way for its subsequent analysis. This process is performed using several C++ tools and algorithms, grouped in different projects. The package which serves as framework for all these projects in LHCb is called Gaudi [99]. The different steps leading the raw detected data to the physics results are summarised next:

1. The amount of data generated by LHC collisions is too high to be directly stored. In this way, it becomes necessary an initial and fast selection which allows to discard most of the events that are not interesting for the physics analysis. This initial selection is called *trigger*. The trigger is considered an online process, in the sense that it takes place almost at the same time that the data is being recorded by the detector. LHCb trigger, which reduces the rate from several millions of events per second to just a few thousand, will be explained in section 3.2.3.1. The group of C++ libraries in which the trigger algorithms are gathered is the Moore [100] project.
2. Data selected by the trigger, pure electronic signals recorded by the different subdetectors, are transformed by different mathematical algorithms in an ensemble of tracks and vertices. Tracks correspond to the charged particles trajectories produced in the collisions (or by decays of other particles) and which go through the detector, and vertices to the point where the proton–proton

collisions (PVs) or the decay of a particle in two or more *daughter* tracks (SVs) took place. In practice, vertices are built from the crossing point of two or more tracks. The reconstruction of tracks and vertices is mainly done with the tracking detectors seen in section 3.2.1.2. The tracking and vertexing at LHCb will be shown in section 3.2.3.2. The information from the PID detectors (section 3.2.1.3) is then added to identify the nature of the tracks, distinguishing, for instance, muons from other particles. The LHCb PID will be seen in section 3.2.3.3. This whole process is called *reconstruction*, and the group of C++ LHCb libraries which contain the relevant tools, Brunel [101]. It is also important to say that, for the correct reconstruction of the LHCb events, the knowledge of the alignment [102, 103] of the detector is crucial. There are specific tools for the alignment of each subdetector. The group of C++ LHCb libraries in which these tools are all gathered is the Alignment project [104]. Figure 3.18 displays a LHCb reconstructed event.

3. Once all the triggered events have been reconstructed, it becomes necessary to separate them according to their physics content. This is done by selecting the different decays using their particular features. For instance $B^0_{(s)} \to \mu^+\mu^-$ or $K^0_S \to \mu^+\mu^-$, with same final state, have different mass and lifetime of the mother, expected separation of the daughters with respect to the PV,... With these selections, the splitting of the data is performed, thus avoiding the different working groups the analysis of the whole triggered dataset. This splitting procedure is called *stripping* in the framework of LHCb, and each of the selections is embedded in a *stripping line*. These kind of selections, performed with the data already on tape, are called offline. The stripped data are then analysed offline, and the the groups of C++ LHCb libraries which contain the relevant tools for stripping and analysis are DaVinci [105] and Erasmus [106].

4. The triggered, reconstructed and stripped dataset has to be then distributed to a series of computing centers spread worldwide. A copy of the raw data from detectors is also saved, with the idea of allowing a later re-reconstruction and stripping once the relevant algorithms have been improved. The process of distributing the data has a double intention. On the one hand, it ensures that the data cannot be lost, regardless any possible technical problem appearing. On the other, it allows physicists an easy access to a distributed computing system of huge power. This distributed system is called the *Grid* [107].

5. The final tool needed to produce physics results is the use of simulated Monte Carlo (MC) data. This is crucial in order to understand several effects and biases created by the group of processes just described, which affects the raw data directly produced in the proton–proton collisions. The production of MC data as similar as possible to the real one is one of the main efforts of all the high energy physics experiments. The C++ libraries at LHCb dedicated to the MC production are contained in Gauss [108]. The MC simulation at LHCb will be briefly explained in section 3.2.3.4.

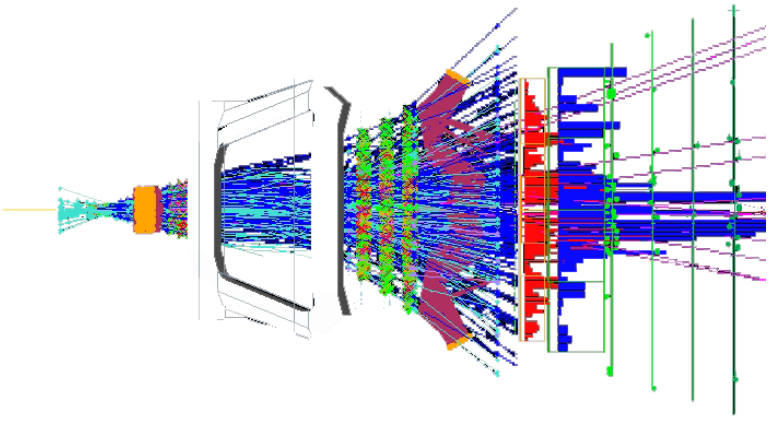

Figure 3.18: Event display of a reconstructed LHCb event, built with Panoramix project [109].

3.2.3.1 Trigger

One of the most important challenges for LHCb is the need of a trigger system capable of reducing the rate of events from proton–proton collisions from the nominal ∼30 MHz (∼15 MHz in 2011) to a maximum rate of ∼3 kHz, which is the maximum permitted by the available resources for long term storage [110, 111, 112]. At the same time, this must be achieved with the minimum possible loss of interesting events for the physics analysis (mainly events in which B and D mesons are present). Two main signatures allow the identification of these kind of events: tracks with high transverse momenta with respect to the beam axis (p_T) and tracks with non-zero impact parameters (IP). The IP is defined as the minimal geometrical distance between a track and the PV of the proton–proton collision.

The LHCb trigger is divided in two different levels [90]. The first level (L0), is a hardware trigger, implemented using custom made electronics to reduce the input rate to a maximum of 1 MHz. At this rate, the whole detector can be read out. The second trigger level (High Level Trigger, HLT) is a C++ application running on an Event Filter Farm (EFF) composed of several thousands CPU nodes. It reduces the L0 output rate to a maximum rate of about 3 kHz. The HLT selected events are then saved on permanent storage. The HLT itself is divided in two parts: HLT1 and HLT2. HLT1 reduces its input rate to about 40 kHz using a partial reconstruction of the data to save computing time. At the HLT2 level, events are reconstructed and selected by a set of inclusive and exclusive algorithms. The reconstruction performed in HLT2 is as similar as possible to the one performed offline, with the limitations coming from computing time requirements.

As a consequence of the stability of the peak luminosity seen in section 3.2.2, the trigger configuration was also kept stable during the overall data taking period, with a L0-output rate exceeding 800 kHz (figure 3.19, left) and a physics trigger HLT output rate of ∼ 3 kHz (figure 3.19, right).

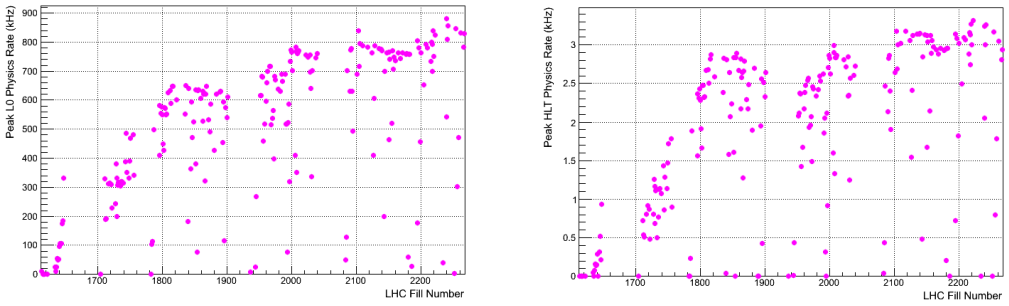

Figure 3.19: L0 trigger output rate (left) and HLT physics trigger output rate (right) as a function of the fill number during 2011.

In LHCb, the trigger configuration information is embedded into an hexadecimal word called *Trigger Configuration Key* (TCK) that identifies the set of trigger decisions, the algorithms run in the trigger and the cuts applied to trigger the events in a given run. Table 3.1 shows the different TCKs used for the 2011 data taking with the corresponding integrated luminosity and magnet polarity.

In spite of all the different TCKs shown, most of the relevant physical trigger lines kept unchanged during the 2011 data taking. Still, all the relevant changes for the analyses between different TKCs will be explained later.

Both L0 and HLT are described next, focusing in their 2011 configuration and performance. The specific trigger issues concerning $B^0_{(s)} \to \mu^+\mu^-$ ($K^0_S \to \mu^+\mu^-$) will be explained in section 5.3 (6.4).

3.2.3.1.1 L0 L0 uses information from selected sub-detectors, which are read-out at the virtual bunch crossing rate (40 MHz): the calorimeters and muon chambers to provide high p_T candidates (hadrons, leptons and photons), and the SPD to provide a fast detection of high occupancy events. L0 positive decisions are sent back to the front-end electronics of all the sub-detectors, which pick-up the pieces of the relevant events from buffers and send them to the EFF.

The L0 decision is taken by the so called L0 Decision Unit (L0DU), based on the following information:

- Calorimeter clusters classified as electrons/positrons, photons, neutral pions or charged hadrons (L0Calo). The identification of these particles is done according to their energy deposits in the different elements of the calorimeter system (see figure 3.13 left).

TCKs	Integrated Luminosity	Magnet Polarity
0x5a0032	39 pb^{-1}	Up
0x5a0032	29 pb^{-1}	Down
0x6d0032	100 pb^{-1}	Down
0x730035	196 pb^{-1}	134 pb^{-1} Up / 62 pb^{-1} Down
0x740036	5.2 pb^{-1}	5.2 pb^{-1} Up
0x760037	298.7 pb^{-1}	107.1 pb^{-1} Up / 191.6 pb^{-1} Down
0x790037	39.3 pb^{-1}	39.3 pb^{-1} Up
0x790038	363.4 pb^{-1}	154 pb^{-1} Up / 209.4 pb^{-1} Down

Table 3.1: TCKs used for the 2011 data taking with the corresponding integrated luminosity and magnet polarity.

- Muon and dimuon candidates (L0Muon).
- Multiplicities at the SPD.

The L0Calo [113] computes the transverse energy deposited in clusters of 4 cells (2×2) using only cells of same sizes. The transverse energy is defined as: $E_T = E \sin\theta$, where E is the energy deposited in one cell and θ is the polar angle with respect to the z axis of the position of the center of the cell.

In the case of L0Muon [114], muon candidates are built searching for straight segments in the four muon stations downstream of the calorimeter, where the occupancy is relatively low. The search is performed under the assumption that the muon tracks roughly point to the interaction point. The segments found are confirmed by looking for a compatible hit in the muon chamber upstream of the calorimeter (M1). The momentum is then estimated by assuming again that the muon tracks originate at the interaction point. The momentum resolution obtained is $\Delta p/p \sim 20\%$.

A positive L0 decision can only be taken when a calorimeter or muon (or dimuon) candidate above the p_T or E_T threshold is found. The threshold values for different types of candidates are shown in table 3.2.

On top of the mentioned p_T and E_T cuts, a Global Event Cut (GEC) is applied to all the events in order to survive L0. The number of hits at the SPD must be below 600 hits in all the cases, except for events in which at least a dimuon candidate is found, for which the maximum number of hits at the SPD is relaxed to 900. For $B^0_{(s)} \to \mu^+\mu^-$ channels, fired at L0 mainly by dimuon triggers, as it will be seen, the gain of this change is at the per mille level (which was verified using a MC simulated sample).

L0 line	
single muon (p_T) [GeV/c]	1.5
dimuon ($\sqrt{p_{T1} \times p_{T2}}$) [GeV/$c$]	1.3
hadron (E_T) [GeV]	3.5
photon (E_T) [GeV]	2.5
electron (E_T) [GeV]	2.5

Table 3.2: Thresholds of the L0 trigger lines used to trigger the 2011 dataset.

3.2.3.1.2 HLT As explained, the HLT is a C++ algorithm run in two sequential parts: HLT1 and HLT2. HLT1 performs a partial reconstruction of the events, while the one at HLT2 is more similar to the offline. With this, computing time is saved while keeping high efficiencies for most of the interesting channels.

HLT1 trigger lines [115, 114] are based on the idea of looking for high transverse momentum tracks with a good track fit quality which are well displaced from all primary interactions. The most relevant HLT1 lines can be classified as follows:

- The simple idea explained is exploited by the *Hlt1TrackAllL0* line, which looks for high p_T and IP tracks regardless the decision at L0.

- In order to increase efficiency for muon decays, for events coming from L0Muon or L0DiMuon and if a track is identified as a muon, the transverse momentum and track quality requirements are relaxed (*Hlt1TrackMuon*).
- Similarly, and again in events from L0Muon or L0DiMuon, if two muons are found so that they form a good vertex, the candidate can be selected based on dimuon mass and on the separation of the dimuon vertex. Depending on the mass, the lines are *Hlt1DiMuonHighMass* and *Hlt1DiMuonLowMass*.
- Finally in events with an L0 Electron or Photon trigger the transverse momentum requirement on the track is slightly relaxed in order to increase efficiency for decays of the type $B \to X\gamma$ (*Hlt1TrackPhoton*).

The thresholds for some of the most relevant cuts in the HLT1 lines just described can be found in table 3.3.

trigger line	
Hlt1AllTrack (p_T [GeV/c] / IP [mm]/ IPχ^2)	1.85/0.1/16
Hlt1TrackMuon (p_T/ IP [mm]/ IPχ^2)	1/0.1/16
Hlt1DiMuonLowMass (p_T[GeV/c]/$M_{\mu\mu}$[GeV/c^2]/IPχ^2)	0.5/1/9
Hlt1DiMuonHighMass (p_T[GeV/c]/$M_{\mu\mu}$[GeV/c^2])	0.5/2.5
Hlt1TrackPhoton (p_T [GeV/c] / IP [mm]/ IPχ^2)	1.2/0.1/16

Table 3.3: HLT1 threshold values used in 2011 run. Only selected trigger cuts are given.

Apart from requiring at least a candidate that survives the mentioned HLT1 cuts, a series of new GEC are also applied. In this way, the number of hits in the OT detector has a maximum value, in order to reject high multiplicity events. This is one of the few cuts that changed depending on the TCK, as the threshold was initially set at 10000 and from TCK 0x6d0032 onwards replaced by 15000. Furthermore, upper cuts in the number of VELO hits (10000) and IT hits (3000) are also imposed.

For the events selected by HLT1, the HLT2 [115] algorithm is applied. At HLT2, a whole event reconstruction, very similar to the one to be applied offline, is performed, to later run the HLT2 selections. These aim for the following signatures:

- Displaced vertices formed by two, three or four tracks.
- Single muons with high p_T and IP
- Pairs of leptons (muons or electrons) with moderate p_T and IP and forming a good vertex.
- A photon cluster in the calorimeter and a pair of tracks forming a secondary vertex, aiming for radiative B decays ($B \to X\gamma$)
- Several exclusive selections of B and D channels.

The only GEC imposed at HLT2 is having 350 or less tracks (built just with VELO).

3.2.3.1.3 Performance The 2011 trigger efficiency on data, calculated for several representative LHCb physics channels, can be found in [113, 114, 115]. In the case of $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$, the trigger efficiencies will be assessed in sections 5.9.5 and 6.7.3.

As a representative example, table 3.4 shows the trigger efficiency on offline selected $B^+ \to J/\psi K^+$ in 2011 data for L0, HLT1 and HLT2.

3.2.3.2 Tracking and vertexing

As explained, one of the key points for the reconstruction of events of events at LHCb is the determination of the trajectories of all the charged particles (tracks) and the position where they were generated (vertices) both if this was a proton–proton collision (PVs) or the decay point of some other particle (SVs). The algorithms that reconstruct tracks seek for the best possible precision in determining

Trigger Level	Efficiency
L0	$(92.5 \pm 0.7)\%$
L0×HLT1	$(86.0 \pm 2.5)\%$
L0×HLT1×HLT2	$(77.0 \pm 2.5)\%$

Table 3.4: Trigger efficiency on offline selected $B^+ \to J/\psi K^+$ in 2011 data for L0, HLT1 and HLT2

the slopes and momenta of the tracks. In the case of vertices, the best resolution is searched for in the 3D position. Finally, the tracking and vertexing algorithms are different online and offline. As already said, the online reconstruction has to be quicker, and this is achieved by reducing the precision with respect to offline.

3.2.3.2.1 Tracking The track reconstruction at LHCb is essentially based on the search for hits all along the tracking detectors likely to have been left by the same particle. This process produces a χ^2 which is related with the "quality" of the track, i.e., with the probability that the track corresponds to a real particle and not to a mixture of hits from different particles (a ghost). The tracking algorithm at LHCb is based on the Kalman Filter [116]. A wide review on LHCb tracking can be found in [117].

Depending on the origin of the hits composing the LHCb tracks, these can be classified as (figure 3.20 left):

- Long tracks: VELO, TT and T–stations. These are the tracks used for most of LHCb analyses. Figure 3.20 right, shows an example of long tracks built by the tracking system.
- Upstream tracks: VELO and TT. Upstream tracks are mainly due to low momentum particles that were bent out of the acceptance by the magnetic field.
- Downstream tracks: TT and T–stations. Decay products from K^0_S and Λ are a typical source of downstream tracks.
- VELO tracks: only VELO. Since the VELO has several modules upstream and downstream of the interaction point, it can reconstruct tracks from backward and large angle particles, used for primary vertex reconstruction.
- T tracks: only hits in the T–stations, typically produced by secondary interactions.

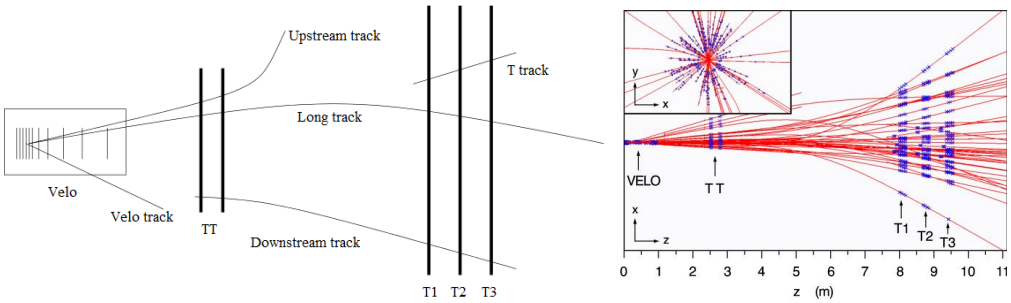

Figure 3.20: Tracking at LHCb. Left, different tracking strategies at LHCb. Right, event display of a series of long tracks with the hits selected by the tracking system.

The efficiency of the LHCb tracking system (referring to both the tracking detector and software) has been assessed in [118] for muons from J/ψ. Figure 3.21 shows this efficiency as a function of momentum and pseudorapidity on 2011 data and MC simulation. The total efficiency has been found to be above 97% for muons of both charges and regardless the polarisation of the magnetic field.

In order to correctly determine the properties of the decays being analysed, one of the most important parameters is the IP of the tracks. The IP resolution will not be constant at LHCb, but dependent on the p_T of the tracks. Figure 3.22 shows the IP_x (IP projected in the x direction) and IP_y (IP projected in the y direction) resolutions as a function of $1/p_T$ in 2011 data, and compares it with the one in MC simulation. As seen in the figure, the behaviour can be correctly described using a linear function. The

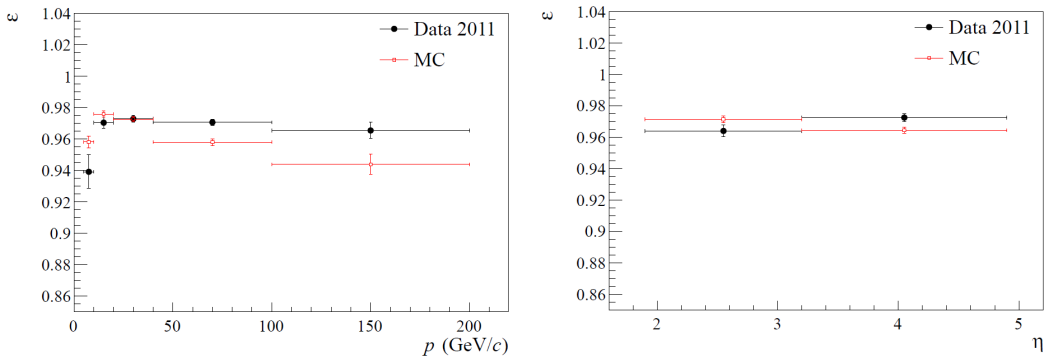

Figure 3.21: Tracking efficiency (tracking detector + tracking software) on muons from J/ψ as a function of momentum (left) and pseudorapidity (right). Black points correspond to 2011 data and red to MC simulation.

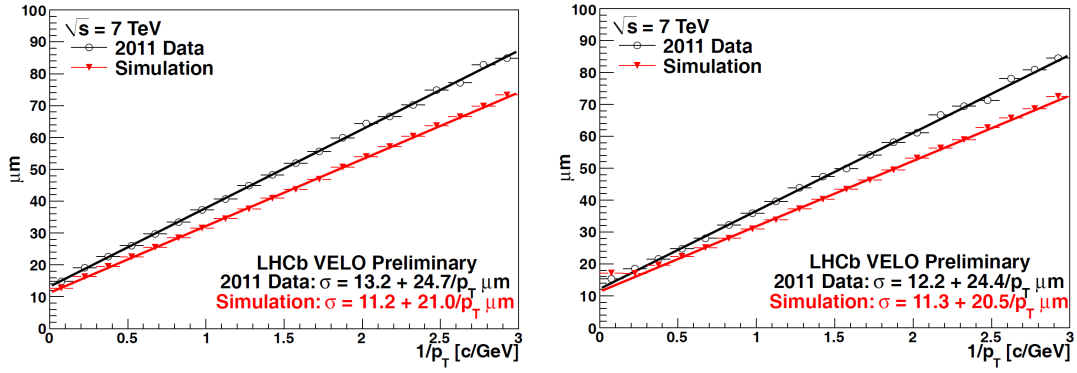

Figure 3.22: IP_x (left) and IP_y (right) resolution as a function of $1/p_T$ in 2011 data (black) and MC (red). A fit to a linear function is also included in both cases.

difference seen between MC and data comes from the description of the material and the interactions of the particles with it in the simulation.

Finally, and as already said, knowing the tracks after traversing the magnetic field allows the determination of its momentum. The momentum resolution is also very important for the physics results, since it is one of the most important parameters entering in the mass resolution (the mass of the decaying particles is obtained from the daughters tracks). Figure 3.23 shows the momentum resolution determined with muons from J/ψ as a function of the momentum itself. The momentum resolution comes defined as $\delta p/p$.

3.2.3.2.2 Vertexing The accurate measurement of production and decay vertices is another important constraint for most of LHCb physics analyses. As explained, LHCb includes a specific detector (VELO), with the idea of giving excellent resolutions in the reconstruction of the positions of PVs and SVs.

The primary vertex resolution is very correlated to the number of tracks used to reconstruct the vertex. The number of tracks making a vertex ranges from 5 (the required minimum) to around 100. More details in PV reconstruction can be found in [119]. Figure 3.24 shows the PV resolution for the x, y and z coordinates in 2011 data and MC simulation as a function of the number of tracks entering in the PV reconstruction.

As for the SV reconstruction, the proper-time resolution achieved (relevant for CP–violation measurements) is at the level of $\sim$ 50 fs, to be compared to typical B lifetimes: $\tau_{B_s} \sim 1500$ fs. Equivalently, the B_s^0 oscillation frequency, measured by LHCb to be $\Delta m_{B_s^0} = (17.63 \pm 0.11) \times \hbar\ ps^{-1}$

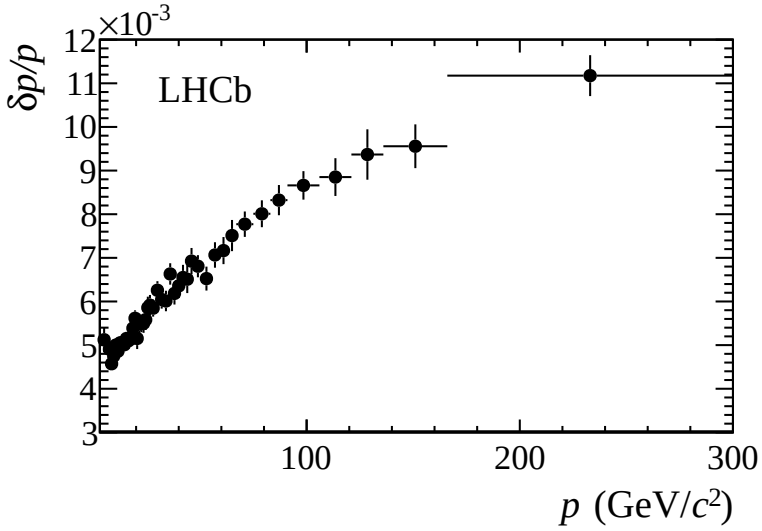

Figure 3.23: Momentum resolution (defined as $\delta p/p$) as a function of p in 2011 data, determined with muons from J/ψ.

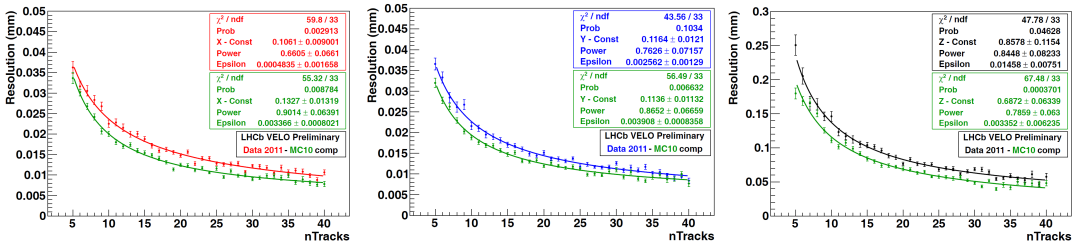

Figure 3.24: PV resolution in x (left), y (center) and z (right) coordinates in 2011 data and MC simulation as a function of the number of tracks entering in the PV reconstruction. The data points are fitted with a three parameter function: $Res = (c1/nTracks^{c2}) + c3$

[120], could have been obtained up to the level of $\sim 50 - 70 \times \hbar\, ps^{-1}$.

3.2.3.3 PID

The particle identification (PID) is the last step of the LHCb events reconstruction. Once all the tracks have been built, the information from the PID detectors (RICH1, RICH2, Calorimeters and Muon System) is added in order to establish hypotheses about the nature of the particles (pion, kaon, proton, muon or electron). Similarly, the energy deposits at the calorimeters may be related to photons, even if these particles cannot be associated to tracks since they are neutral.

The LHCb PID from the different subdetectors is usually combined in a common likelihood, in order to maximise the efficiency and minimise the misidentification rate. However, the identification of each particle is typically dominated by one of the subdetectors. The PID algorithms for RICH, Calo and Muon System, together with the performances in 2011 data are presented next.

3.2.3.3.1 RICH The RICH PID is based on the difference in the expected Cherenkov angle (angle of the radiated photons) between protons, kaons and pions. The low mass difference between pions and muons ($\sim$139 vs. $\sim$105 MeV/c^2) makes the Cherenkov angle very similar in both these particles, so RICH PID is not efficient separating them. Figure 3.25 shows the Cherenkov angle as a function of momentum for muons, pions, kaons and protons.

Using the RICH information, likelihoods can be built for tracks in order to establish the probabilities of different particle hypotheses [90]. As an example, figure 3.26 left demonstrates the kaon efficiency (kaons identified as kaons) and pion misidentification (pions misidentified as kaons) fraction on 2011

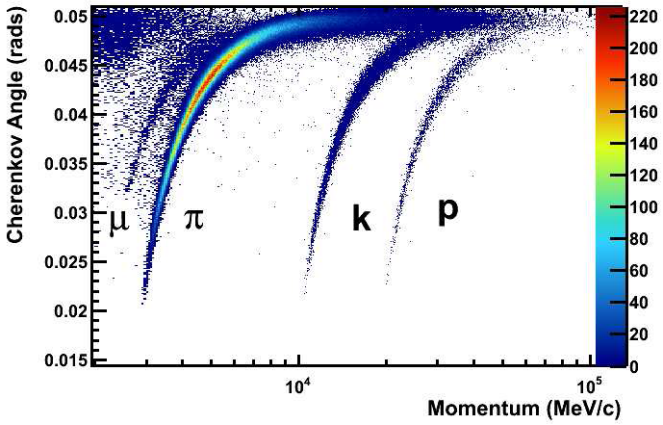

Figure 3.25: Reconstructed Cherenkov angle as a function of track momentum in the C_4F_{10} radiator.

data, as a function of momentum, achievable with two different cuts in the log difference of these likelihoods: DLL(K–π). Similarly, figure 3.26 right shows the proton efficiency and pion misidentification fraction as a function of momentum again with two different cuts in the DLL(p–π).

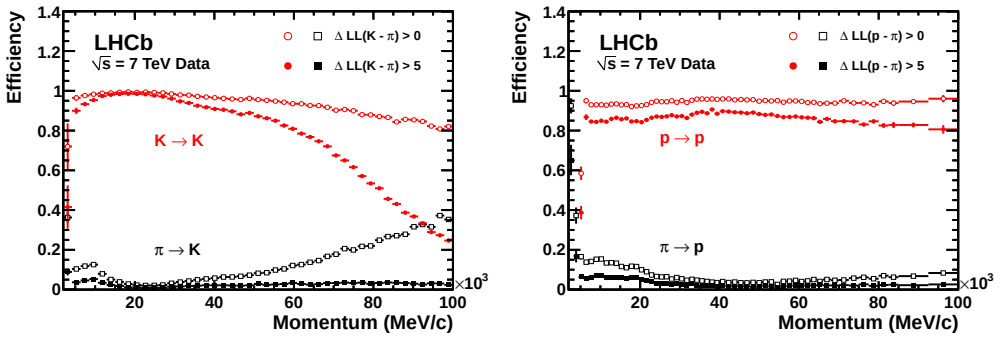

Figure 3.26: RICH PID performance on 2011 data. Left, kaon efficiency and pion misidentification fraction, as a function of momentum, with two different cuts in the DLL(K–π). Right, proton efficiency and pion misidentification fraction as a function of momentum with two different cuts in the DLL(p–π).

3.2.3.3.2 Calo CALO PID uses the energy deposits in the different regions (see figure 3.13 left) of the LHCb calorimeter to identify the particle having gone through the calorimeter. It is important to say that, as muons are minimum ionizing particles (MIPs), they are capable of going through the detector losing a characteristic energy on it which can also can help on its identification.

A review on the PID calorimeter algorithms can be found in [121]. Similarly to the RICH case, but using here the energy deposited in the different calorimeter elements, these algorithms use the information to obtain likelihoods for the different particle hypotheses. Results on the Calo PID performance of electrons (photons) with 2011 data can be found in [122] ([123]). As an example, figure 3.27 shows the efficiency and misidentification rate for electrons in 2011 data as function of momentum and for different DLL(e–h) cuts, where the h stands for hadrons.

3.2.3.3.3 Muon MuonID uses the fact that muons are the only particles (except for hadrons punch-through, which are rather unlikely) able to go through the calorimeter and hit the muon system stations. For any reconstructed track, a field of interest at the muon stations is built and muon hits are searched for. The identity and location of these hits allows the calculation of a muon likelihood that is

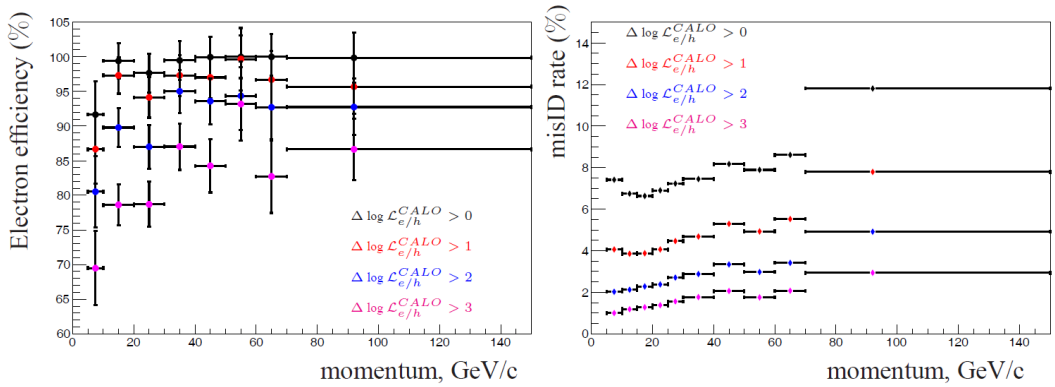

Figure 3.27: Electron PID performance vs. momentum on 2011 data for different cuts in the DLL(e–h). Left, efficiency. Right, misidentification rate.

later used to discriminate between muons and other particles. The muonID algorithms are described in [124]. An alternative method has also been introduced in [125].

It is important to say that the main source of misidentified muons are pions and hadrons decaying in flight to a combination $\mu^-\bar{\nu}_\mu$ (or $\mu^+\nu_\mu$). In this case, and depending on the kink of the muon with respect to the initial hadron, the hadron decaying in flight may become almost indistinguishable from a real muon. This is very important for the $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$ analyses.

The performance of the muon identification in 2011 data has been assessed in [126]. Figure 3.28 shows the muon PID efficiency and misidentification rates for protons and pions vs. momentum for different cuts in the DLL(μ–p).

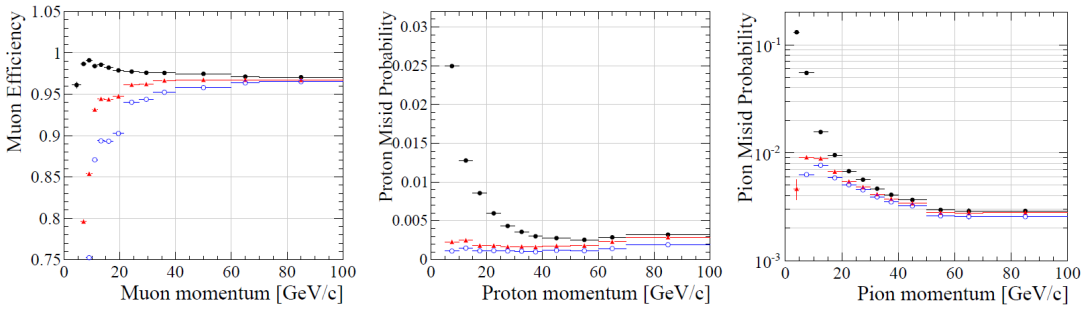

Figure 3.28: Muon PID performance vs. momentum on 2011 data for different cuts in the DLL(μ–p). Left, efficiency on muons. Center, misidentification rate for protons. Right, misidentification rate for pions.

3.2.3.4 MC

High energy physics analyses also have as a key ingredient the simulation of data with the MC method. The importance of MC simulated data arises for several reasons. On the one hand, it permits to train the tools that are later used on data, but with the advantage of having a full knowledge in advance of the process being studied. On the other, it is useful to obtain certain parameters needed for the physics results but very difficult to be extracted from data itself. Both these features will be seen and used with various particular examples in the $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$ analyses.

As mentioned, the MC data used at LHCb [127] are generated using the simulation application Gauss, which is a collection of libraries for physics simulation based on Gaudi and with specialised algorithms and tools for generators (PYTHIA [128], EvtGen [129], ...) and detector simulation (Geant4 [130]).

The MC events generated in LHCb can be classified as follows:

- Minimum Bias: Keep all events generated by PYTHIA: elastic, diffractive, inelastic.
- Inclusive: Extract events generated by PYTHIA with at least one b or c hadron in 400 mrad with respect to the LHCb z axis. If all of these hadrons have $p_z < 0$, flip the whole event.
- Signal: Extract events generated by PYTHIA containing at least one specific particle in 400 mrad. Again, if the candidate has p_z <0, flip the whole event. In the case of b hadrons and to speed up generation, if the interaction contains the b, repeat the hadronisation process of PYTHIA until the interaction contains the correct particle.

These techniques will be used to generate the $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$ decays (among others), and will produce useful samples of the signal searched for.

4

Search for rare decays at LHCb

The "rare decay" tag is usually given to modes in which the branching fractions are suppressed in the SM and the rates or other observables can be calculated experimentally precisely enough to give sensitivity to NP.

Rare decays analysed at LHCb are all FCNC, so they cannot occur at tree level within the SM framework. In this way, the access to NP is through new virtual particles entering in the loops (see, for instance, figure 2.9). Therefore, this is an indirect search of NP, with the advantage of accessing higher energy scales than in direct searches.

These searches also have several common experimental features:

- Use of control channels to avoid dependence on simulation.
- Normalisation to channels with similar geometry/trigger to convert an observed number of events in a BR, resulting in smaller systematic errors.
- Combination of geometrical properties with multivariate analysis (MVA [131]) tools, to classify the events according to their similarity to signal.
- Need for good particle ID (particularly muon) and low pion/kaon misID.
- Perform blind analyses, so that the signal region is not looked at until all the issues are completely understood.

The goal of this chapter is reviewing the most important LHCb results in rare decays using 2011 dataset. The rare decays at LHCb are typically subdivided in electroweak penguin transitions, radiative decays and very rare decays.

It is important to recall that among the 6 key channels for LHCb, identified in [89], 3 are rare decays. These 3 channels are representative of each of the subgroups just mentioned. The 3 key modes are $B^0 \to K^{*0}\mu^+\mu^-$ (electroweak penguin), $B_s^0 \to \phi\gamma$ (radiative decay) and $B_s^0 \to \mu^+\mu^-$ (very rare decay).

Both $B_{(s)}^0 \to \mu^+\mu^-$ and $K_S^0 \to \mu^+\mu^-$ (which also enters in the very rare decays group) will not be explained here as they will have specific chapters.

4.1 Electroweak penguin transitions $b \to d\mu^+\mu^-$ and $b \to s\mu^+\mu^-$

4.1.1 $B^0 \to K^{*0}\mu^+\mu^-$

$B^0 \to K^{*0}\mu^+\mu^-$ is a $b \to s\mu^+\mu^-$ FCNC decay, mediated by electroweak penguin diagrams in the SM. An example of two of these diagrams can be found in figure 4.1.

The study of the angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ is a very sensitive test of NP. This angular distribution can be described by three angles defined in the B rest frame (θ_L, θ_K and ϕ) and it varies with the dimuon mass (q^2). Figure 4.2 shows the definition of the three mentioned angles.

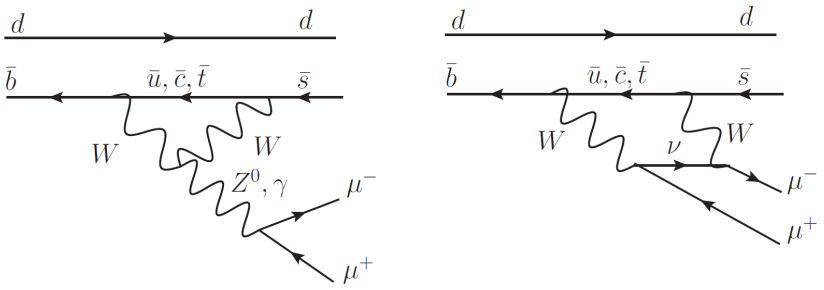

Figure 4.1: Examples of $B^0 \to K^{*0}\mu^+\mu^-$ Feynman diagrams in SM

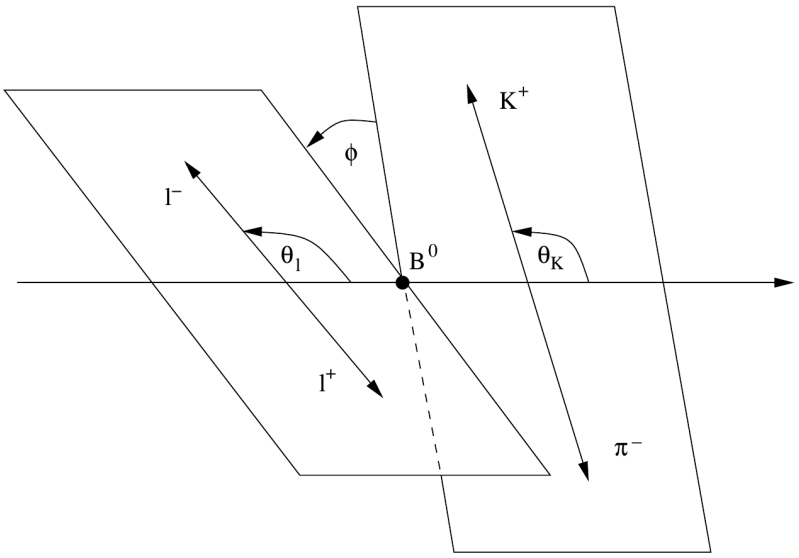

Figure 4.2: Definition of the θ_L, θ_K and ϕ angles, used to describe the $B^0 \to K^{*0}\mu^+\mu^-$ angular distribution.

For the analysis, the angular distribution of $B^0 \to K^{*0}\mu^+\mu^-$ is parameterised in terms of 4 theoretically clean observables, with different NP sensitivities:

- A_{FB}, the forward-backward asymmetry;
- F_L, the fraction of K^{*0} longitudinal polarisation;
- S_3, the asymmetry in the K^{*0} transverse polarisation [132];
- S_9, CP average of A_{IM} [133].

The analysis at LHCb with the 1 fb^{-1} 2011 data [134], has been performed by suppressing the combinatorial background using a Boosted Decision Tree (BDT) [135], and rejecting the peaking backgrounds with different PID requirements. The resonant contributions from $B^0 \to J/\psi K^{*0}$ and $B^0 \to \Psi(2S)K^{*0}$ have also been vetoed. With this, ∼900 signal events have been obtained, whose mass spectra can be seen in figure 4.3.

For the angular analysis, a correction (taken from simulation), is applied to correct the angular bias arising from the detector acceptance. For each bin in q^2, a 4D fit is performed, using the information from the θ_L, θ_K, ϕ angles and the B^0 mass. The results obtained are shown in figure 4.4 as a function of q^2, together with the SM predictions. The values obtained are world best, and compatible with the theory expectancy.

Another observable that has been obtained in the $B^0 \to K^{*0}\mu^+\mu^-$ analysis is the zero crossing point of the A_{FB} asymmetry, which is the point where A_{FB} changes sign and which corresponds to a well defined point in q^2. This zero crossing point has been obtained in data using an unbinned-counting technique, so that the q^2 distributions of forward and backward going candidates are extracted in an

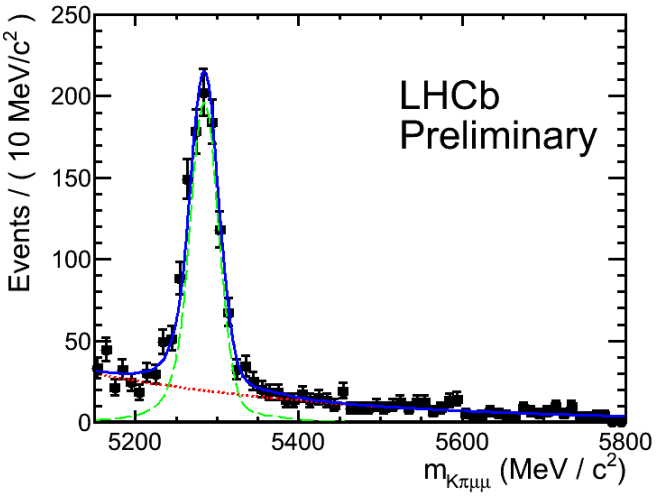

Figure 4.3: $B^0 \to K^{*0}\mu^+\mu^-$ mass spectra with the ∼900 events obtained from the LHCb 2011 data. A fit is also included to obtain the total yield of $B^0 \to K^{*0}\mu^+\mu^-$. For details on the fit see [134].

independent way by performing unbinned maximum–likelihood fits to the $K^+\pi^-\mu^+\mu^-$ invariant mass and q^2 distributions. Then, both distributions are properly normalised and subtracted in order to extract the zero crossing point. The A_{FB} obtained by the unbinned counting can be compared in figure 4.5 left with the result of a simple counting method and with the theory prediction. The zero-crossing point is measured to be $q^2 = (4.9^{+1.1}_{-1.3})$ GeV^2/c^4 , in agreement with the SM prediction.

The final measurement to be mentioned from the work in [134] is the differential $\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$ as a function of q^2. This differential BR has been previously measured by other experiments and has accurate SM predictions. Figure 4.5 right shows the LHCb results using 2011 data and the comparison with the theory prediction. Both results are again in agreement.

It is important to say that, even when compatible with SM, all these results produce stringent bounds in the NP phase space (see for instance [136]).

4.1.2 Isospin analysis in $B \to K^{*0}\mu^+\mu^-$

The combination of pairs of channels, such as $B^0 \to K^0(\to K^0_S)\mu^+\mu^- \leftrightarrow B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^{*0}\mu^+\mu^- \leftrightarrow B^+ \to K^{*+}(K^0_S\pi^+)\mu^+\mu^-$ allows the construction of clean observables from ratios or asymmetries, in which some theoretical errors cancel. These 4 channels are all FCNC, and their SM contribution is dominated by diagrams similar to those of figure 4.1. An example of the combined observable just mentioned is the CP averaged isospin asymmetry (A_I), defined as:

$$\begin{aligned} A_I &= \frac{\Gamma(B^0 \to K^{(*)0}\mu^+\mu^-) - \Gamma(B^+ \to K^{(*)+}\mu^+\mu^-)}{\Gamma(B^0 \to K^{(*)0}\mu^+\mu^-) + \Gamma(B^+ \to K^{(*)+}\mu^+\mu^-)} \\ &= \frac{\mathcal{B}(B^0 \to K^{(*)0}\mu^+\mu^-) - \frac{\tau_0}{\tau_+}\mathcal{B}(B^+ \to K^{(*)+}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{(*)0}\mu^+\mu^-) + \frac{\tau_0}{\tau_+}\mathcal{B}(B^+ \to K^{(*)+}\mu^+\mu^-)}, \end{aligned} \tag{4.1}$$

where $\Gamma(B \to f)$ and $\mathcal{B}(B \to f)$ are the partial width and branching fraction of the $B \to f$ decay and $\frac{\tau_0}{\tau_+}$ is the ratio of the lifetimes of the B^0 and B^+ mesons.

LHCb has measured A_I as a function of the mass of the dimuon pair (q^2) using the data taken during 2011 [137]. For that, the BR of the 4 mentioned channels has been obtained as a function of q^2. These included two modes with K^0_S, for which the reconstruction efficiency is low, due to the long K^0_S lifetime. In order to increase the statistics, the K^0_S built with a pair of downstream tracks (see section 3.2.3.2.1) have been included and analysed independently. The BRs of $B^0 \to K^0_S\mu^+\mu^-$, $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^{*0}\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ have been obtained by normalising to the J/ψ modes ($B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$). As in the $B^0 \to K^{*0}\mu^+\mu^-$ analysis, both modes were

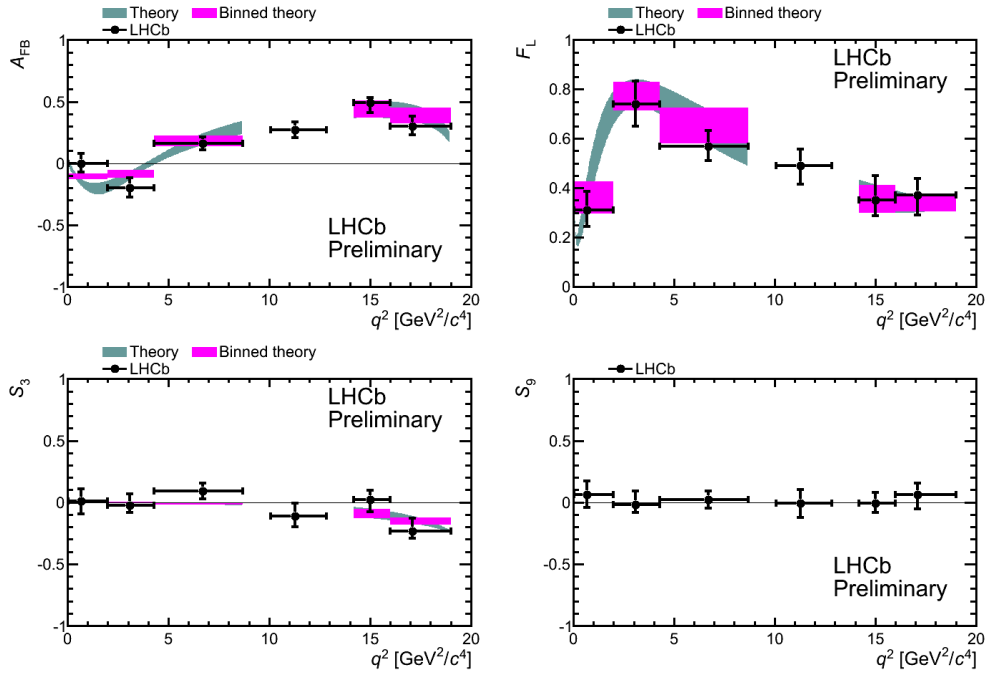

Figure 4.4: $B^0 \to K^{*0}\mu^+\mu^-$ angular analysis observables as a function of the dimuon mass (q^2). From top to bottom and from left to right: A_{FB}, F_L, S_3 and S_9. The SM predictions are also shown when available (green and magenta bars).

vetoed from the selection. Figure 4.6 shows the mass spectra obtained for the four modes. In the case of $B^0 \to K^0_S\mu^+\mu^-$, this is a first world observation, with a significance of 5.7 σ.

The A_I asymmetry obtained by LHCb can be seen in figure 4.7 both for the $B \to K\mu^+\mu^-$ and $B \to K^{*0}\mu^+\mu^-$ modes. In the case of the $B \to K^{*0}\mu^+\mu^-$, the result obtained is compatible with 0, as expected in the SM. In the case of $B \to K\mu^+\mu^-$, A_I is significantly below 0 (more than 4 σ). In this case, even if a SM theory result is not present, the theoretical prediction is expected to be close to 0.

4.1.3 $B^+ \to \pi^+\mu^+\mu^-$

The $B^+ \to \pi^+\mu^+\mu^-$ is a $b \to d\mu^+\mu^-$ transition very suppressed in the SM with respect to $B^+ \to K^+\mu^+\mu^-$, due to the difference in the corresponding CKM matrix elements, $|V_{td}|$ and $|V_{ts}|$.

LHCb has observed $25.3^{+6.7}_{-6.4}$ candidates in the 2011 data sample [138], corresponding to a significance of 5.2 σ. Figure 4.8 shows the mass distribution of these events. Normalising to $B^+ \to J/\psi K^+$, a $\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-) = (2.4 \pm 0.6\ (\text{stat.}) \pm 0.2\ (\text{syst.})) \times 10^{-8}$ has been measured. This BR is consistent with the SM prediction. This is the rarest B decay ever observed, and the first $b \to d\mu^+\mu^-$ transition.

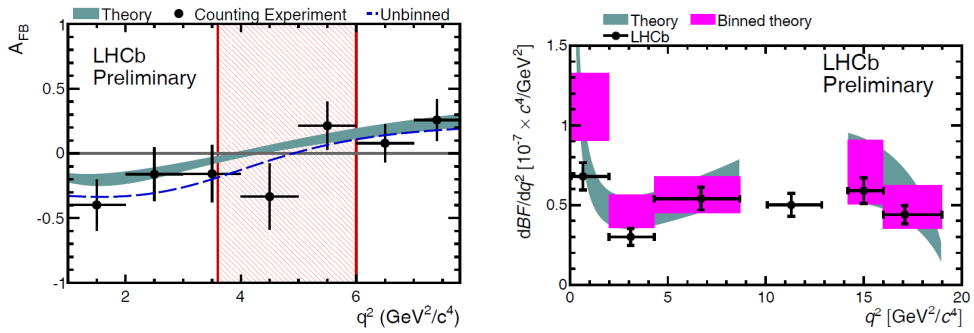

Figure 4.5: Extra measurements performed in the $B^0 \to K^{*0}\mu^+\mu^-$ analysis. Left, extraction of the zero crossing point of A_{FB}, including the unbinned method and the simple counting method (see text for details), the theory prediction and the 68% CL region in the data result (red hatched region). Right, differential $\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$ as a function of q^2, including the SM prediction.

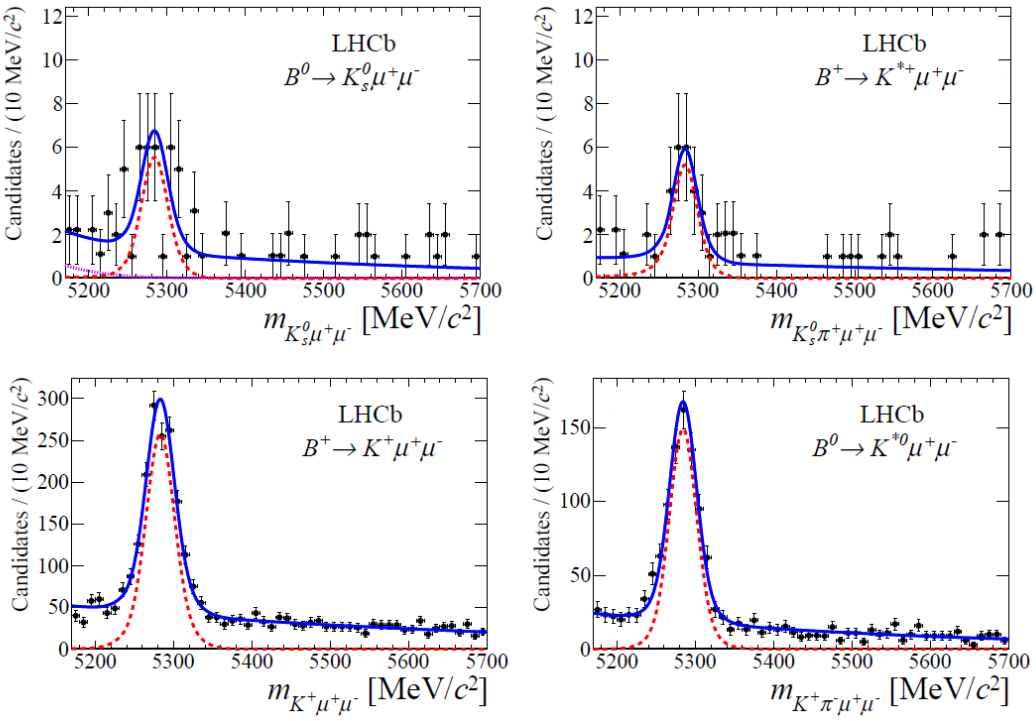

Figure 4.6: Mass spectra of (from top to bottom and from left to right) $B^0 \to K^0_S\mu^+\mu^-$, $B^+ \to K^{*+}\mu^+\mu^-$, $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ obtained with 2011 data. In the modes with K^0_S, the downstream–downstream category is not shown. Fits are also performed to obtain the total yield of each mode. For details on the fit see [137].

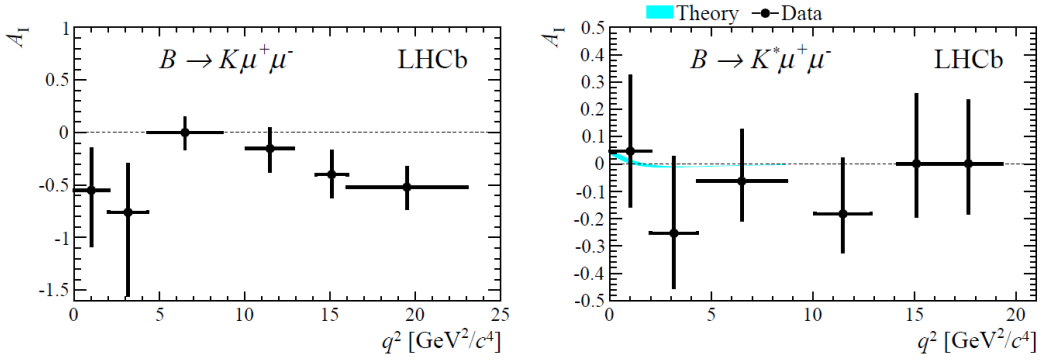

Figure 4.7: A_I obtained by LHCb with 2011 data. Left, $B \to K\mu^+\mu^-$. Right, $B \to K^{*0}\mu^+\mu^-$. The SM prediction is shown where available.

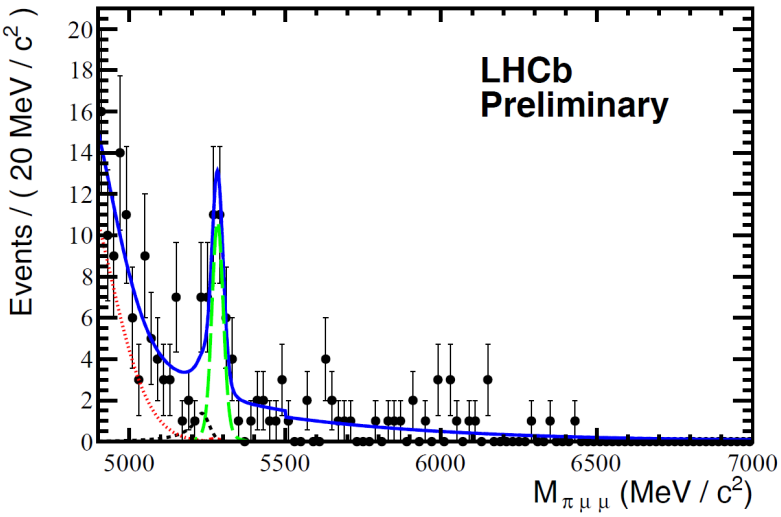

Figure 4.8: Mass spectrum of the selected $B^+ \to \pi^+\mu^+\mu^-$ candidates obtained by LHCb with 2011 data. A fit is also performed to obtain the $B^+ \to \pi^+\mu^+\mu^-$ total yield. For details on the fit see [138].

4.2 Radiative decays

In the SM, the radiative decays of B mesons take place at leading order through $b \to s\gamma$ diagrams, in one-loop electromagnetic penguin transitions. These transitions are dominated by a virtual intermediate top quark coupling to a W boson. Figure 4.9 shows an example of one of the mentioned transitions, for the case of the $B^0 \to K^{*0}\gamma$ decay. The radiative b decays can be affected by NP, since there are extensions of the SM predicting other contributions that can introduce sizeable effects on the dynamics of the radiative transitions. The direct CP asymmetry (A_{CP}) in $b \to s\gamma$ transitions is one of the observables sensitive to NP effects.

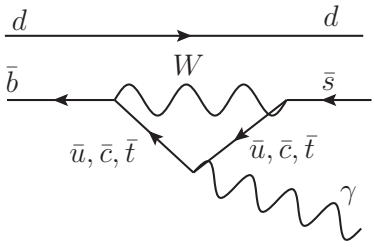

Figure 4.9: Example of one-loop electromagnetic penguin transition contributing to $B^0 \to K^{*0}\gamma$ decay.

The CP asymmetry in radiative decays has been measured by LHCb with the 2011 data using the $B^0 \to K^{*0}\gamma$ decay mode [139]. The analysis implies a simultaneous fit for the raw asymmetry between B^0 and $\bar{B}^0$ decays. Figure 4.10 shows the mass spectra for the $B^0 \to K^{*0}\gamma$ and $\bar{B}^0 \to \bar{K}^{*0}\gamma$ decays. The measured raw asymmetry is simply related to the relative yield between both modes. In order to extract the final result, the B^0 production and the $K\pi$ detection asymmetries need to be corrected. With all this, A_{CP} is measured to be $A_{CP}(B^0 \to K^{*0}\gamma) = -0.008 \pm 0.017$ (stat.) ± 0.009 (syst.). This result is consistent with the SM prediction.

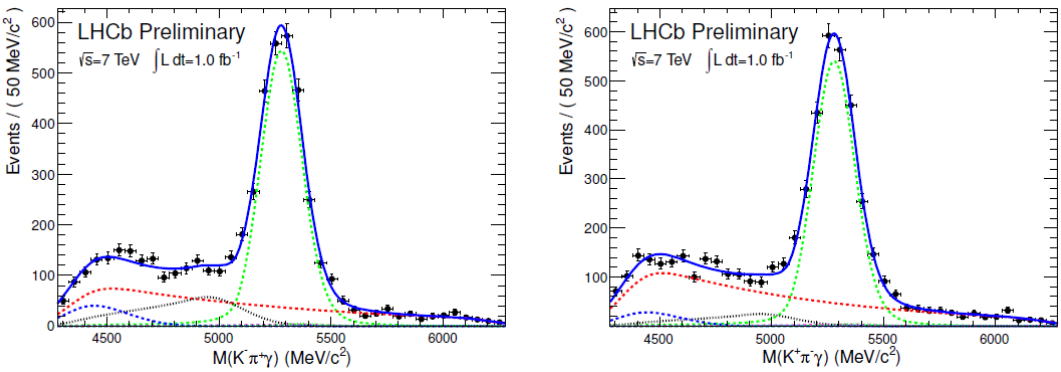

Figure 4.10: Simultaneous fit of the 2011 LHCb data for $\bar{B}^0 \to \bar{K}^{*0}\gamma$ (left) and $B^0 \to K^{*0}\gamma$ (right). For details on the fit see [139].

4.3 Very rare decays

4.3.1 $\tau^- \to \mu^+\mu^-\mu^-$

$\tau^- \to \mu^+\mu^-\mu^-$ is a lepton flavour violating (LFV) mode, so therefore extremely suppressed within the SM. In the SM, $\tau^- \to \mu^+\mu^-\mu^-$ can only be mediated by diagrams with neutrino oscillation (figure 4.11 left), leading to a BR far beyond the current experimental sensitivity scope (see references in [140]).

However, this decay is enhanced in several NP scenarios, in which new particles in the loops allow LFV modes. An example are the LH models, with diagrams as the one in figure figure 4.11 right.

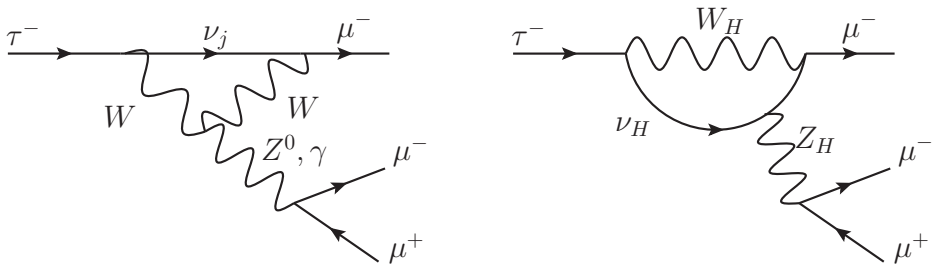

Figure 4.11: Feynman diagrams contributing to $\tau^- \to \mu^+\mu^-\mu^-$ decay. Left, SM. Right, LH models.

LHCb has profited the large production of τ within its acceptance (mainly from B and D mesons) to establish upper limits on the $\mathcal{B}(\tau^- \to \mu^+\mu^-\mu^-)$ using the 2011 dataset [140]. The analysis uses a loose selection to later classify all the candidates according to three variables: a BDT which combines geometrical and kinetical information, a PID BDT which uses the PID information from different subdetectors, and the $\mu^+\mu^-\mu^-$ invariant mass. The normalisation has been done to the $D_s^- \to \phi(\mu^+\mu^-)\pi^-$ channel, taking into account the fraction of τ coming from D_s, the $\mathcal{B}(D_s^- \to \phi(\mu^+\mu^-)\pi^-)$, and the $\mathcal{B}(D_s^- \to \tau^-\bar{\nu}_\tau)$. Apart from the combinatorial background, the $D_s^- \to \eta(\mu^+\mu^-\gamma)\mu^-\bar{\nu}_\mu$ decay has also been studied and taken into account. Figure 4.12 shows the events observed in the four merged bins that contain the highest and second highest signal probability. No signal excess has been observed in the τ mass region. With all this, the upper limit in the $\mathcal{B}(\tau^- \to \mu^+\mu^-\mu^-)$ has been established to be $\mathcal{B}(\tau^- \to \mu^+\mu^-\mu^-) < 6.3\ (7.8) \times 10^{-8}$ at 90(95)% CL. This limit is competitive with the current world best, established by the Belle collaboration [5].

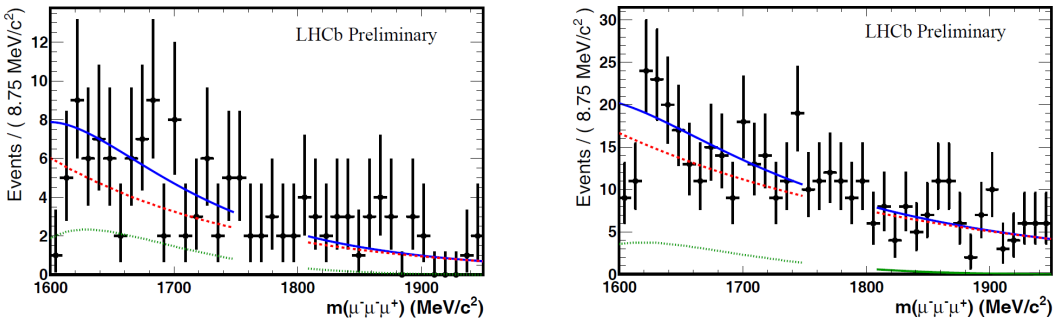

Figure 4.12: $\mu^+\mu^-\mu^-$ observed candidates during 2011 by LHCb in the four merged bins that contain the highest (left) and second highest (right) signal probability. A fit is included accounting for combinatorial exponential and $D_s^- \to \eta(\mu^+\mu^-\gamma)\mu^-\bar{\nu}_\mu$ contributions. For more details in the fit see [140].

4.3.2 $B^0_{(s)} \to \mu^+\mu^-\mu^+\mu^-$

B^0 and B^0_s to four muons decays are strongly suppressed in the SM. The largest contribution to these decays comes with the double two muons resonance $B^0_s \to J/\psi(\mu^+\mu^-)\phi(\mu^+\mu^-)$ with an expected BR at the level of $\sim 2\times10^{-8}$ [5]. Nevertheless, the non–resonant decay is also possible in the SM with diagrams mediated by virtual photons, $B^0_{(s)} \to \mu^+\mu^-\gamma^*(\mu^+\mu^-)$, although the expected BR is several orders of magnitude smaller than the resonant one (see references in [141]). In spite of this, the advantage of the non–resonant mode is that it can be sensitive to NP, through new particles decaying into $\mu^+\mu^-$ pairs (see again references in [141]).

LHCb has been the first experiment ever to search for the non–resonant decays of $B^0_{(s)}$ mesons to four muons, using the data taken during 2011 [141]. In order to get rid of the mentioned resonant contribution, the selection included vetoes in the J/ψ and ϕ masses of the dimuon pairs. The events found by LHCb in the signal region were compatible with the background expectations, and limits on the BRs have been obtained in this analysis by normalising to $B^0 \to J/\psi K^{*0}$. Figure 4.13 shows the invariant mass spectra of the $\mu^+\mu^-\mu^+\mu^-$ candidates in the B^0 and B^0_s mass regions. The limits extracted, $\mathcal{B}(B^0 \to \mu^+\mu^-\mu^+\mu^-) < 5.4 \times 10^{-9}$ and $\mathcal{B}(B^0_s \to \mu^+\mu^-\mu^+\mu^-) < 1.3 \times 10^{-8}$ (at 95% CL), are, as explained, the world first on these decays.

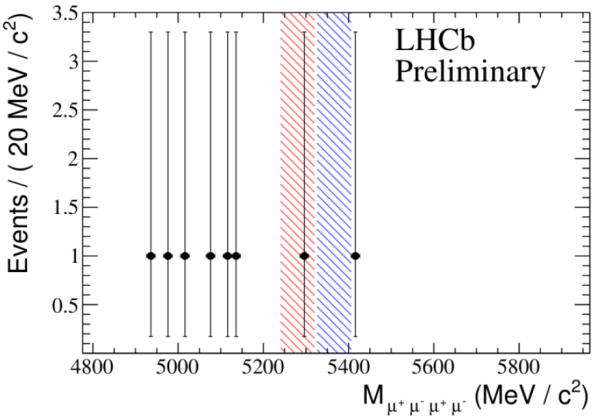

Figure 4.13: Invariant mass spectra of the non resonant $\mu^+\mu^-\mu^+\mu^-$ selected candidates using 2011 LHCb dataset. The red (blue) area shows the B^0 (B^0_s) mass region.

4.3.3 Searches for Majorana Neutrinos

Charged B decays to final states containing hadrons plus a $\mu^\pm\mu^\pm$ pair are LFV, so forbidden or extremely suppressed in the SM. However, they can occur if mediated by a potential Majorana Neutrino. In particular, modes such as $B^- \to D^{*+}\mu^-\mu^-$ and $B^- \to D^+\mu^-\mu^-$ can arise from the presence of a virtual Majorana neutrino, while $B^- \to D_s^+\mu^-\mu^-$, $B^- \to \pi^+\mu^-\mu^-$ and $B^- \to D^0\pi^+\mu^-\mu^-$ can be mediated by an on–shell Majorana neutrino. Figure 4.14 shows examples of Feynman diagrams of the mentioned decays, all mediated by Majorana neutrinos.

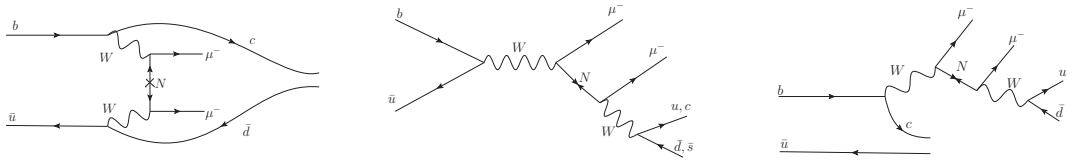

Figure 4.14: Feynman diagrams contributing to LFV decays, all mediated by an intermediate Majorana neutrino (N). Left, $B^- \to D^{(*)+}\mu^-\mu^-$. Center, $B^- \to D_s^+(\pi^+)\mu^-\mu^-$. Right $B^- \to D^0\pi^+\mu^-\mu^-$.

Using 0.37 fb^{-1} of 2011 and 0.04 fb^{-1} of 2010 data, LHCb has searched for B charged decays mediated by Majorana neutrinos, using the 5 particular modes explained before [142]. The analysis has been performed using $B^- \to J/\psi K^-$ ($B^- \to \Psi(2S)K^-$) to normalise the modes with 3 (5) tracks in the final state. Since the search has not produced any signal excess, upper limits on the BR of these modes have been set. This is shown in table 4.1. In the case of the last three modes, as the mediating Majorana neutrino appears on-shell, limits have been established as a function of the neutrino mass. In particular, the limits in the $\pi^+\mu^-\mu^-$ final state have been used to establish neutrino mass dependent upper limits on the coupling $|V_{\mu 4}|$ of a heavy Majorana neutrino to a muon and a virtual W. This is shown in figure 4.15.

Channel	Upper limit on BR
$B^- \to D^{*+}\mu^-\mu^-$	2.4×10^{-6}
$B^- \to D^+\mu^-\mu^-$	6.9×10^{-7}
$B^- \to D_s^+\mu^-\mu^-$	5.8×10^{-7}
$B^- \to \pi^+\mu^-\mu^-$	1.3×10^{-8}
$B^- \to D^0\pi^+\mu^-\mu^-$	1.5×10^{-6}

Table 4.1

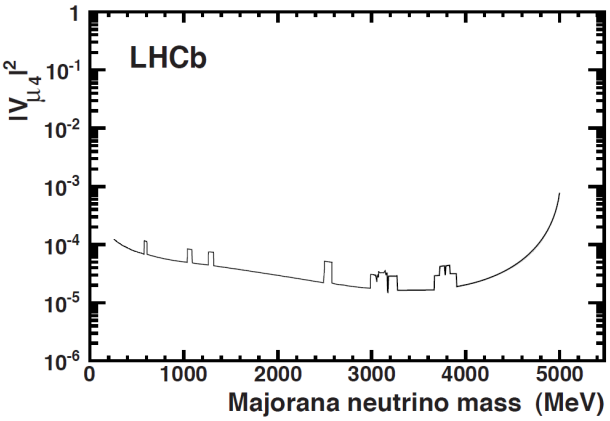

Figure 4.15: Upper limits on $|V_{\mu 4}|^2$ at 95% CL as a function of the Majorana neutrino mass. These limits have been obtained with 0.41 fb^{-1} of 2010 and 2011 LHCb data, using the $B^- \to \pi^+\mu^-\mu^-$ decay mode.

4.3.4 $D^0 \to \mu^+\mu^-$

Similarly to $K_S^0 \to \mu^+\mu^-$, $D^0 \to \mu^+\mu^-$ decay is dominated by LD in SM, but has a SD component through which NP can enhance its BR well above the SM prediction. Following [143], the SM prediction, in the range $\sim 10^{-13} - 10^{-12}$, goes up, e.g, to $\sim 10^{-9}$ in RPV models. Figure 4.16 shows an example of a diagram contributing to $D^0 \to \mu^+\mu^-$ in these models.

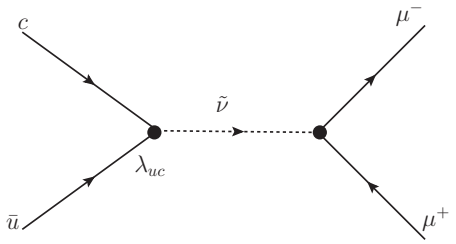

Figure 4.16: Example of RPV diagram contributing to $D^0 \to \mu^+\mu^-$.

LHCb has searched for $D^0 \to \mu^+\mu^-$ using 0.9 fb^{-1} of 2011 data [143]. In order to get a cleaner sample, only the tagged subsample ($D^{*+} \to D^0\pi^+$) has been used. Apart from the combinatorial background (for whose reduction a BDT has been designed), this decay suffers from the presence of $D^0 \to h^+h'^-$ (where h is either a kaon or a pion) in which both hadrons are misidentified as muons. In order to properly establish the presence of the signal or extract a limit, this double misID has to be studied in a quantitative way, estimating the yield in the D^0 mass region. Finally, the normalisation of the $D^0 \to \mu^+\mu^-$ has been done to $D^0 \to \pi^+\pi^-$, where again the D^0 are restricted to those coming from D^{*+}. Finally, the yields have been extracted using 2D fits, in which both the D^0 mass and the difference $M(D^{*+}) - M(D^0)$ are used. With this, LHCb has not found any signal excess, but it has

established a world best upper limit in the BR of this channel, $\mathcal{B}(D^0 \to \mu^+\mu^-) < 1.3 \times 10^{-8}$ at 95% CL. Figure 4.17 shows the $\mu^+\mu^-$ mass spectra in two particular $M(D^{*+}) - M(D^0)$ ranges, together with the fit performed.

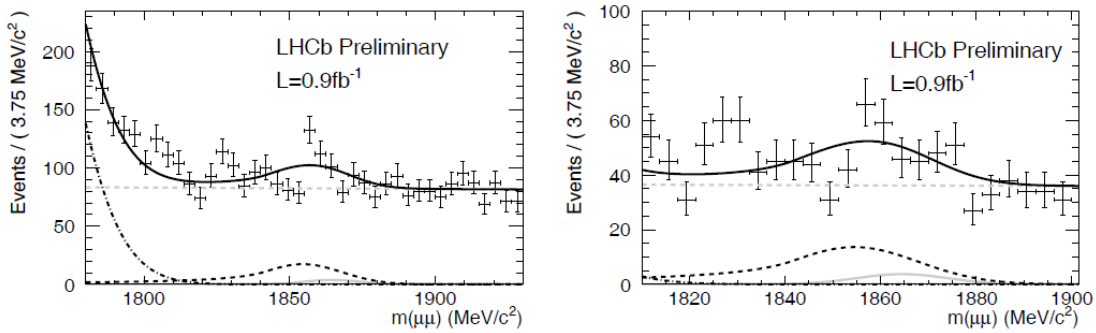

Figure 4.17: $\mu^+\mu^-$ mass spectra in two particular $M(D^{*+}) - M(D^0)$ ranges from 0.9 fb^{-1} of LHCb 2011 data. Left, 142-149 MeV/c^2. Right, 144-147 MeV/c^2. The fits performed to obtain the yields are also shown. For details in the fit, see [143].

5

$B^0_{(s)} \to \mu^+\mu^-$

5.1 Introduction

The $B^0_{(s)} \to \mu^+\mu^-$ decays are very important for the search for NP in the flavour sector (see section 2.2.3). Because of this, they were identified as a "key" channels for LHCb [89], as already said in the previous chapter. The SM prediction for the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ ($\mathcal{B}[B^0_s \to \mu^+\mu^-]$) was shown in equation 2.10 (equation 2.11) to be $[1.1 \pm 0.1] \times 10^{-10}$ ($[3.2 \pm 0.3] \times 10^{-9}$).

$B^0_{(s)} \to \mu^+\mu^-$ decays have been widely searched for by various experiments (including LHCb), in last years. The best published limits from the Tevatron are $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 5.1 \times 10^{-8}$ at 95% confidence level (CL) by the D0 collaboration using 6.1 fb^{-1} of data [144], and $\mathcal{B}(B^0 \to \mu^+\mu^-) < 6.0 \times 10^{-9}$ at 95% C.L. by the CDF collaboration using 6.9 fb^{-1} of data [145]. In the same dataset the CDF collaboration observes an excess of $B^0_s \to \mu^+\mu^-$ candidates compatible with $\mathcal{B}(B^0_s \to \mu^+\mu^-) = (1.8^{+1.1}_{-0.9}) \times 10^{-8}$ and with an upper limit of $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 4.0 \times 10^{-8}$ at 95% C.L. . The CMS collaboration has recently published $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 1.9 \times 10^{-8}$ at 95% C.L. and $\mathcal{B}(B^0 \to \mu^+\mu^-) < 4.6 \times 10^{-9}$ at 95% C.L. using 1.14 fb^{-1} of data [146]. The LHCb collaboration has published the limits [147] $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 1.6 \times 10^{-8}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.6 \times 10^{-9}$ at 95% C.L. based on about 370 pb^{-1} of integrated luminosity collected in the first half of the 2011 run. A combination of these results with the LHCb limits obtained with the 2010 dataset [148] leads to $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 1.4 \times 10^{-8}$ and $\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.2 \times 10^{-9}$ at 95% C.L. . For the case of $B^0_s \to \mu^+\mu^-$, figure 5.1 summarises the experimental situation at the end of 2011.

This chapter presents a search for $B^0_{(s)} \to \mu^+\mu^-$ using the data recorded by LHCb in 2011, which correspond to an integrated luminosity of ~ 1.02 fb^{-1}.

Assuming the branching fractions predicted by the SM, and using the $b\bar{b}$ cross-section measured by LHCb of $\sigma_{b\bar{b}} = 75 \pm 14\,\mu$b [150] (obtained in the pseudorapidity interval $2 < \eta < 6$ and integrated over all transverse momenta), approximately 11 $B^0_s \to \mu^+\mu^-$ and 1 $B^0 \to \mu^+\mu^-$ events are expected to be reconstructed and selected in the analysed sample.

The analysis proceeds as follows. First a very efficient selection removes most of the background while keeping very high efficiency for signals. The number of observed events is then compared to the number of expected signal and background events in bins of two independent variables, the invariant mass and the output of a multi-variate discriminant, the BDT [135].

The probability for a signal or background event to have a given value of the BDT output is extracted from data using $B^0_{(s)} \to h^+h'^-$ candidates (where $h^{(\prime)}$ can be a pion or a kaon) as signal and sideband $B^0_{(s)} \to \mu^+\mu^-$ candidates as combinatorial background.

The invariant mass line shape of the signals is described by a Crystal Ball function [151] whose parameters are extracted from data using control samples. The central values of the masses are obtained from $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ samples. The B^0_s and B^0 mass resolutions are estimated by interpolating the resolution measured on di-muon resonances ($J/\psi, \psi(2S)$ and $\Upsilon(1S, 2S, 3S)$). The

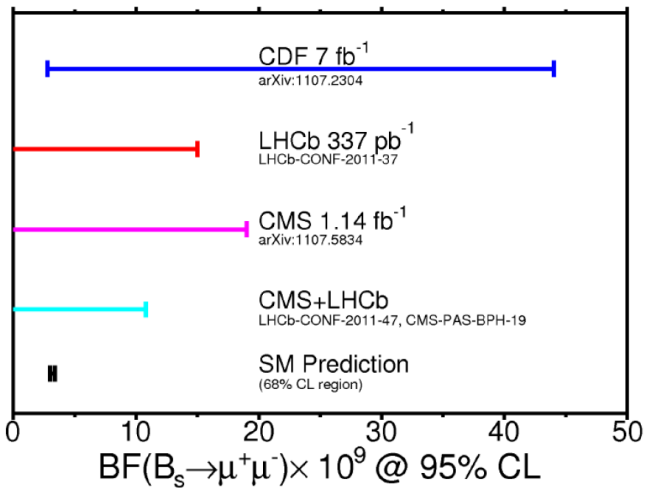

Figure 5.1: Experimental status in the search for $B^0_s \to \mu^+\mu^-$ at the end of 2011. The SM prediction is also included. Extracted from [149].

parameters of the radiative tails have been extracted from a fit to the mass distribution of $B^0_s \to \mu^+\mu^-$ simulated events where the resolution has been smeared in order to reproduce the measured one. The central values and the resolutions of the B^0 and B^0_s masses are used to define the search windows.

The number of expected signal events, for a given branching fraction hypothesis, is obtained by normalising to channels with known branching fractions: $B^+ \to J/\psi\, K^+$, $B^0_s \to J/\psi\, \phi$ and $B^0 \to K^+\pi^-$. These channels are selected in a way as similar as possible to the signals in order to minimise the systematic uncertainty related to the different phase space accessible to each final state.

The BDT output and invariant mass distributions for combinatorial background events in the signal regions are obtained using fits of the mass distribution of events in the mass sidebands in bins of the BDT output.

There is also a peaking background contribution composed by $B^0_{(s)} \to h^+h'^-$ in which both hadrons are misidentified as muons. The expected mass and BDT distributions of this background is obtained with the help of MC and the misidentification rates measured with kaons and pions from $D^0 \to K\pi$ decay. The total expected peaking background is derived from the total number of $B^0_{(s)} \to h^+h'^-$ in data.

The two–dimensional space formed by the invariant mass and the BDT output is binned. For each bin, the number of candidates observed in the data are counted, and the expected number of signal and background events computed. The compatibility of the observed distribution of events in all bins with the distribution expected for a given branching fraction hypothesis is computed using the $\mathrm{CL_s}$ method [152, 153], which allows a given hypothesis to be excluded at a given confidence level.

In order to avoid any unconscious bias, the mass region $m_{\mu\mu} = [m(B^0) - 60\,\mathrm{MeV}/c^2, m(B^0_s) + 60\,\mathrm{MeV}/c^2]$ has been blinded until the completion of the analysis.

5.2 Data and Monte Carlo Samples

5.2.1 Data Sample

As explained, the $B^0_{(s)} \to \mu^+\mu^-$ analysis is performed using the 2011 LHCb dataset, which includes $\sim$ 1.02 fb^{-1} of integrated luminosity taken at $\sqrt{s} = 7$ TeV.

The data belong to the Reco12-Stripping17 campaign, having been reconstructed with Brunel v41r1, the Condition Data Base [154] (condDB) head-20111111 and the Detector Data Base (DDDb) [155] head-20110914 and analysed with DaVinci v29r2.

5.2.2 Monte Carlo samples

Most of the Monte Carlo samples used in the $B^0_{(s)} \to \mu^+\mu^-$ analysis belong to the Monte Carlo production MC10, sim01, except for a few exclusive background samples, where MC11, sim05 was chosen. The Monte Carlo samples are listed in table 5.1, and a brief reminder of the generation/reconstruction conditions is reported below.

The *pp* interactions have been simulated assuming a beam energy of 3.5 TeV, an average number of interactions per crossing $\nu = 2.5$, which corresponds to an average number of visible interactions per crossing $\mu = 1.75$, and a $\beta^* = 3.5$ m.

The software with which the MC10 samples have been generated belongs to Gauss v39r0, based on LHCb v26r3 [156], Pythia 6_418.2, PHOTOS 215.2 [157], LHAPDF 5.3.1 [158], HepMC 1.26, and GEANT4 v91r3.p03. Then, they have been reconstructed using Boole [159] v21r9 and Brunel v37r8p5. The software used to generate MC11 samples can be found in section 6.2.2.

As explained in section 3.2.3.4, and to save CPU time, cuts are applied at the generation level to enforce particles of interest to be between 10 mrad and 400 mrad. The particles that fulfill this criteria are: the two muons of the signal samples, the two muons of $B^0_s \to J/\psi\phi$, $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$ control channels and the two hadrons of $B^0_{(s)} \to h^+h'^-$ decays. For the $b\bar{b} \to \mu\mu X$ background sample, two muons are required to be in the acceptance regardless of their charge. The generator level efficiencies due to the acceptance cuts are shown in table 5.2, taken from [160]. The total number of $b\bar{b} \to \mu\mu X$ events corresponds to ~ 570 pb^{-1} of integrated luminosity, assuming the $b\bar{b}$ cross section measured by LHCb, $\sigma_{b\bar{b}} = 288 \pm 4 \pm 48$ μb [161].

The trigger has been emulated in *pass-through mode* with the configuration that has been more extensively used during the 2011 data taking (see section 3.2.3.1) TCK 0x2d0032. In addition, the same stripping selections (stripping17, see section 5.4) used for data have been applied.

A smearing procedure to the tracks in simulated events has been applied in order to reproduce the impact parameter resolution measured in data. This will be discussed in section 5.9.4.1.

Channel	Monte Carlo production	Magnet polarity	Events (processed)	Events (reco & strip.)
Signal:				
$B^0_s \to \mu^+\mu^-$	MC10sim01	Down / Up	∼500 k	∼ 180 k
Background:				
$b\bar{b} \to \mu\mu$,p >3 GeV/c, $M_{\mu\mu} > 4.7\,\text{GeV}/c^2$	MC10sim01	Down	50 M	57 k
$b\bar{b} \to \mu\mu$,p >3 GeV/c, $M_{\mu\mu} > 4.7\,\text{GeV}/c^2$	MC10sim01	Up	50 M	57 k
Other specific backgrounds	MC10sim01/ MC11sim05	Down/ Up		
Control/normalisation channels:				
$B^0_s \to J/\psi\phi$	MC10sim01	Down/Up	–	200 k
$B^+ \to J/\psi K^+$	MC10sim01	Down/Up	–	365 k
$B^0 \to J/\psi K^{*0}$	MC10sim01	Down/Up	–	200 k
$B^0 \to K^+\pi^-$	MC10sim01	Down/Up	–	300 k

Table 5.1: Monte Carlo samples used in the $B^0_{(s)} \to \mu^+\mu^-$ analysis. The Monte Carlo production version, the magnet polarity, the number of processed events and the number of reconstructed and stripped events are shown in the second, third, fourth and fifth columns, respectively.

	$B^0_{(s)} \to \mu^+\mu^-$	$B^+ \to J/\psi K^+$	$B^0_s \to J/\psi\,\phi$	$B^0_{(s)} \to h^+h'^-$	$B^0 \to J/\psi K^{*0}$	$b\bar{b} \to \mu\mu X$
$\epsilon_{\rm gen}$	17.5%	15.3%	16.0%	17.7%	14.8%	6.1×10^{-4}

Table 5.2: Generator level efficiency $\epsilon_{\rm gen}$ for signals, control channels and background channel. The uncertainty on the quoted numbers is below 0.1 %.

5.3 Trigger

The LHCb trigger has been described in section 3.2.3.1, where some of the of the trigger lines relevant for the $B^0_{(s)} \to \mu^+\mu^-$ analysis were explained, particularly for L0 and HLT1. The lines responsible for triggering the channels needed for the $B^0_{(s)} \to \mu^+\mu^-$ analysis are briefly reported next.

At L0 level, the signal and the control channels with muons or $J/\psi(\mu\mu)$ in the final state are mostly triggered by the L0-μ and L0-diμ decisions, while the $B^0_{(s)} \to h^+h'^-$ channels are triggered mostly by the L0-hadron decision.

As far as the HLT1 is concerned, table 3.3 includes all the relevant lines, with the Hlt1AllL0Track triggering the hadronic control channels and all the muon lines the rest (including the signal).

Finally, for the HLT2 the small changes in the configuration of the trigger during 2011 play a role. In this way, the channels containing J/ψ in the final state have been triggered in the first 370 pb^{-1} by the *Hlt2DiMuonJPsi* trigger, which requires two identified muons in a 120 MeV/c^2 mass window around the J/Ψ mass. This line was then prescaled by a factor of five for the remaining 650 pb^{-1}, and the channels containing J/ψ in the final state were mostly triggered by the *Hlt2DiMuonDetached* line, where a soft cut on the distance between the secondary and primary vertex divided by its error ($DLS > 3$) was added to the previous selection criteria. Therefore the trigger efficiency map used for the current analysis has been obtained from the J/ψ of the exclusive $B^+ \to J/\psi K^+$ decay (section 5.9.5.1).

The $B^0_{(s)} \to \mu^+\mu^-$ signal candidates are selected with an equivalent trigger selection, *Hlt2DiMuonBmm* but requiring their invariant mass to be above 4.7 GeV/c^2. Some extra J/ψ and $B^0_{(s)} \to \mu^+\mu^-$ events are selected by the *topological muon* lines [115], where only one muon is requested and combined with a second track.

The $B^0_{(s)} \to h^+h'^-$ hadronic channels used for calibration and/or normalisation are selected by the *Hlt2Topo2Body* trigger, a generic $B \to hh + X$ selection and by the exclusive *Hlt2B2hh* selection [115].

5.4 Selection

The initial selection for the $B^0_{(s)} \to \mu^+\mu^-$ analysis has a double intention:

1. reduce the data size to a manageable level by keeping the efficiency on the signal as high as possible; the separation between signal and background is then left to the main discriminant, the Boosted Decision Tree (BDT) described in section 5.7;
2. deal in similar way with signal and control/normalisation channels in order to minimise the systematic uncertainties in the computation of the normalisation factors.

Table 5.3 shows the cuts used to select $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h'^-$ inclusive decays. The $B^0_{(s)} \to h^+h'^-$ inclusive sample is selected similarly to the $B^0_{(s)} \to \mu^+\mu^-$ signals (apart from the muonID requirement), as it is the main control sample for the extraction of the BDT and the invariant mass PDFs (probability density functions) from data. Moreover, the exclusive $B^0 \to K^+\pi^-$ channel is one of the three normalisation channels. It should also be said that the KL variable in the table refers to the Kullback-Leibler (KL) distance [162, 163, 164], which is used to suppress duplicated tracks created by the reconstruction. Finally, as already explained, the signal regions ($\pm$60 MeV/c^2) around the measured B^0 and B^0_s masses have been blinded until the completion of the analysis.

Among the selection listed in table 5.3, a series of fiducial cuts to reject unphysical signal candidates can be found. These include cuts in the p, p_T of the muons and in the proper time of the B candidates, and have an efficiency of $\epsilon = 99.9$ % on Monte Carlo $B^0_s \to \mu^+\mu^-$ signal events. To show it, and also the background rejection, the distributions of the p and p_T of one of the two muons and the proper time of the B candidate can be found in figure 5.2 for $B^0_s \to \mu^+\mu^-$ Monte Carlo events, $b\bar{b} \to \mu\mu X$

	Cut	value
		$B_s^0 \to \mu^+\mu^-$ and $B_{(s)}^0 \to h^+h'^-$
μ / h	track $\chi^2/ndof$	< 4
	IPχ^2	>25
	p_T	> 0.25 and < 40 GeV/c
	p	< 500 GeV/c
	KL	< 5000
μ only	ISMUON	true
$B_{(s)}$	$\lvert M_{hh} - M(B_{(s)})\rvert$	<600 MeV/c^2
	DOCA	<0.3 mm
	vertex χ^2	<9
	VDS	> 15
	IPχ^2	< 25
	t	$< 9 \cdot \tau(B_s^0)$
	BDTS	> 0.05
	DLL$(K-\pi)$	< 10
	DLL$(\mu-\pi)$	> -5

Table 5.3: Selection for $B_{(s)}^0 \to \mu^+\mu^-$ and $B_{(s)}^0 \to h^+h'^-$ channels.

Monte Carlo events and background events in data mass sidebands ([4.9–5.0] GeV/c^2 and [5.433 – 6.0] GeV/c^2).

The $B_{(s)}^0 \to \mu^+\mu^-$ and $B_{(s)}^0 \to h^+h'^-$ selections also include a cut on the transverse momentum of the B candidate $p_T(B) > 500$ MeV/c. This cut has the goal of vetoing the exclusive dimuon production from the process $pp \to p\mu^+\mu^-p$, in which the interchange of two virtual photons between the protons produces the pair of muons. As in these events the final state protons remain undetected in the beam pipe, their only experimental signature are the muons. Moreover, since the dimuon invariant mass reaches the B^0 and B_s^0 mass regions, these dimuon events become a dangerous background in the search for $B_{(s)}^0 \to \mu^+\mu^-$ if they are associated to a primary vertex from a second interaction in the same bunch crossing. In these cases, they can appear very signal like as they can have a large separation and a good pointing to the wrongly associated PV and the tracks in the final state are very isolated. Fortunately, the p_T spectrum of these events is very soft, so they can be efficiently removed with the mentioned cut in p_T. For a more detailed discussion see [165].

Furthermore, and in order to reduce efficiently the 2011 dataset, dominated by the background of the $B_{(s)}^0 \to h^+h'^-$ inclusive sample, a multivariate Boosted Decision Tree discriminant for the selection (BDTS) has been implemented and a loose cut on it has been applied. The definition of the BDTS discriminant is discussed in section 5.4.1.

The muonID cuts, applied to the $B_{(s)}^0 \to \mu^+\mu^-$ candidates, will be analysed in section 5.5.

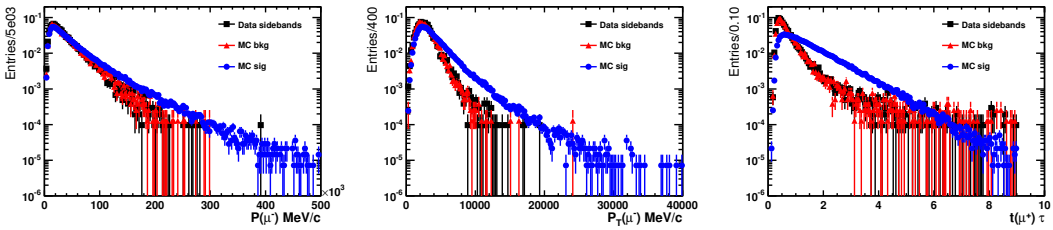

Figure 5.2: Muon momentum (left), transverse momentum (center) and B proper time (right) distributions for $B_s^0 \to \mu^+\mu^-$ Monte Carlo events (blue curve), $b\bar{b} \to \mu\mu X$ Monte Carlo events (red curve) and $B_s^0 \to \mu^+\mu^-$ candidates in data sidebands (black curve).

For the $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ normalisation channels, the additional $K^\pm$ candidates are required to pass the same track quality cut ($\chi^2/ndof < 4$). A IPχ^2 cut is applied on the $K^\pm$ of the

$B^+ \to J/\psi K^+$ decay (IPχ^2 >25) and on the two kaons of the $B^0_s \to J/\psi\phi$ decay (IPχ^2 >4,4) (table 5.4).

For $B^0_s \to J/\psi\phi$ candidates, the KK invariant mass is required to be within $\pm$ 10 MeV/c^2 of the ϕ PDG mass and the $K^\pm$ have to be identified as kaons by the RICH system ($DLL(K-\pi) > 0$). For these normalisation channels, the B candidate IPχ^2 and vertex separation (VDS) requirements are the same as those for the signal selection (IPχ^2 <25, VDS>15).

$B^+ \to J/\psi K^+$			$B^0_s \to J/\psi\phi$		
	Cut	value		cut	value
μ	track $\chi^2/ndof$	< 4	μ	track $\chi^2/ndof$	< 4
	IPχ^2	>25		IPχ^2	>25
	ISMUON	true		ISMUON	true
J/ψ	DOCA	<0.3 mm	J/ψ	DOCA	<0.3 mm
	vertex χ^2	<9		vertex χ^2	<9
	Δm	<60 MeV		Δm	<60 MeV
	VDS	>15		VDS	>15
$K^\pm$	track $\chi^2/ndof$	<5	$K^\pm$	track $\chi^2/ndof$	<5
	IPχ^2	>25		IPχ^2 $[K^+, K^-]$	$[> 4, > 4]$
				$DLL(K-\pi)$ $[K^+, K^-]$	> 0
			ϕ	Δm	<10 MeV
				IPχ^2	> 25
B_u	IPχ^2	<25	B_s	IPχ^2	<25
	Δm	<100 MeV		Δm	<100 MeV
	vertex χ^2	< 45		vertex χ^2	<75
	BDTS	> 0.05		BDTS	>0.05

Table 5.4: Selections for $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$ normalisation channels.

5.4.1 The BDTS discriminant

The BDTS discriminant has been found to be the most efficient way to further reduce the background sample while keeping high efficiency on the signal. The variables entering in the BDTS are:

- the impact parameter (IP(B)) and impact parameter χ^2 (IPχ^2(B)) of the B candidate;
- the χ^2 of the secondary vertex (VCHI2);
- the angle between the direction of the momentum of the B candidate and the direction defined by the difference between the secondary and the primary vertex (DIRA);
- the minimum distance between the two daughter tracks (DOCA);
- the minimum impact parameter of the muons with respect any primary vertex (minIP(μ)).

Only the variables DOCA and IP(B) are in common with the BDT used in the limit computation (see section 5.7). As it will be seen later, and in opposition to the BDT case, this choice of variables also allows the calculation of the BDTS in the normalisation channels.

The primary vertex used to compute the IP(B), IPχ^2(B) and DIRA variables is identified by extrapolating to the beam axis the line defined by the momentum of the B candidate and taking the vertex with the closest distance with respect to the intersection of this line with the beam axis. This vertex is refitted after the removal of the tracks belonging to the B decay chain.

The BDTS distribution for $B^0_s \to \mu^+\mu^-$ Monte Carlo sample (left) and for dimuon sidebands candidates (right) in data are shown in figure 5.3: as for the BDT used for the limit computation, the distribution has been kept uniform between 0 and 1 for signal being, as a consequence, peaked at zero for background. The background distribution agrees nicely with the expected distribution from $b\bar{b} \to \mu\mu X$ Monte Carlo events.

In figure 5.4 the rejection versus efficiency curve is shown for the $B^0_s \to \mu^+\mu^-$ Monte Carlo and $B^0_{(s)} \to h^+h'^-$ sidebands samples: a signal efficiency of $\sim$ 93% can be obtained for a background rejection of $\sim$ 70% applying a BDTS cut BDTS>0.05.

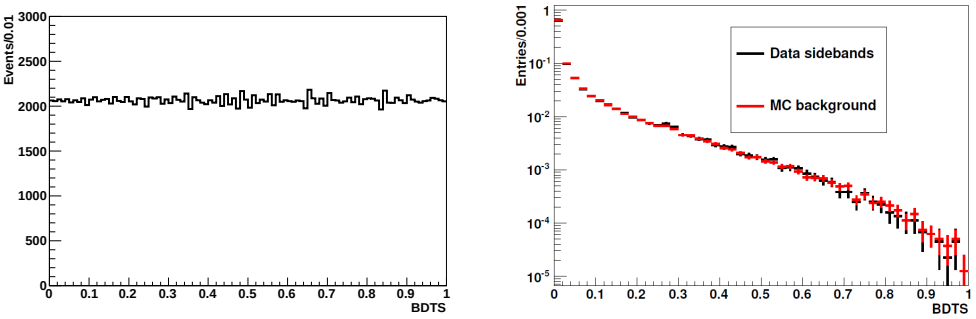

Figure 5.3: BDTS distributions. Left: $B_s^0 \to \mu^+\mu^-$ Monte Carlo events; right: $B_s^0 \to \mu^+\mu^-$ sidebands events (black markers) and $b\bar{b} \to \mu\mu X$ MC events (red markers).

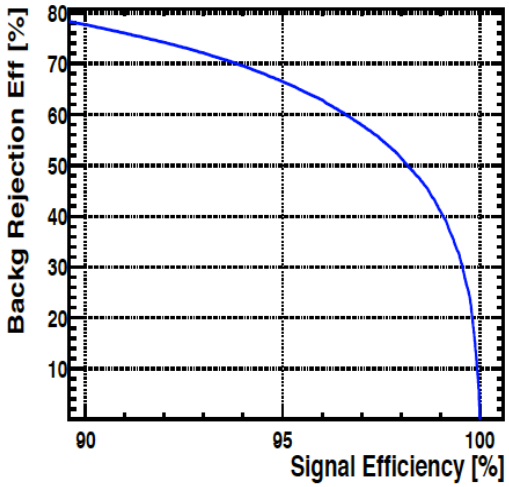

Figure 5.4: Rejection versus efficiency curve for BDTS built for $B_s^0 \to \mu^+\mu^-$ Monte Carlo and $B_{(s)}^0 \to h^+h'^-$ sidebands events.

In order to minimise the systematic uncertainty in the normalisation factors, the same BDTS cut is also applied to the three normalisation channels. For the $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ modes the χ^2 of the secondary vertex is substituted by the χ^2 of the J/ψ vertex, the flight distance is computed between the J/ψ vertex and the primary vertex and the DOCA is computed between the two muons from the J/ψ decay. In this way, the distributions of all the variables but minIP, are very similar for $B_s^0 \to \mu^+\mu^-$, $B_{(s)}^0 \to h^+h'^-$, $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$, resulting in a very similar efficiency for signal and normalisation channels.

In figure 5.5 the BDTS distribution for $B_s^0 \to \mu^+\mu^-$ (top), $B^+ \to J/\psi K^+$ (center) and $B_s^0 \to J/\psi\phi$ (bottom) are shown using Monte Carlo samples non-smeared[1] (left), smeared (center) and over-smeared (right). In table 5.5 the efficiencies of different BDTS cuts on $B_s^0 \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ are shown for smeared Monte Carlo samples. The efficiency ratios for unsmeared, smeared and oversmeared $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ simulated samples are shown in figure 5.6, always being 1 within 0.4%.

The BDTS cut has been chosen by looking at the combined performance of the BDTS and the BDT described in section 5.7. The rejection versus efficiency curves for the BDT as a function of different

[1]the "smearing" is randomisation of the MC track parameters (such as the IP) to make them look as similar as possible to data, see section 5.9.4.1.

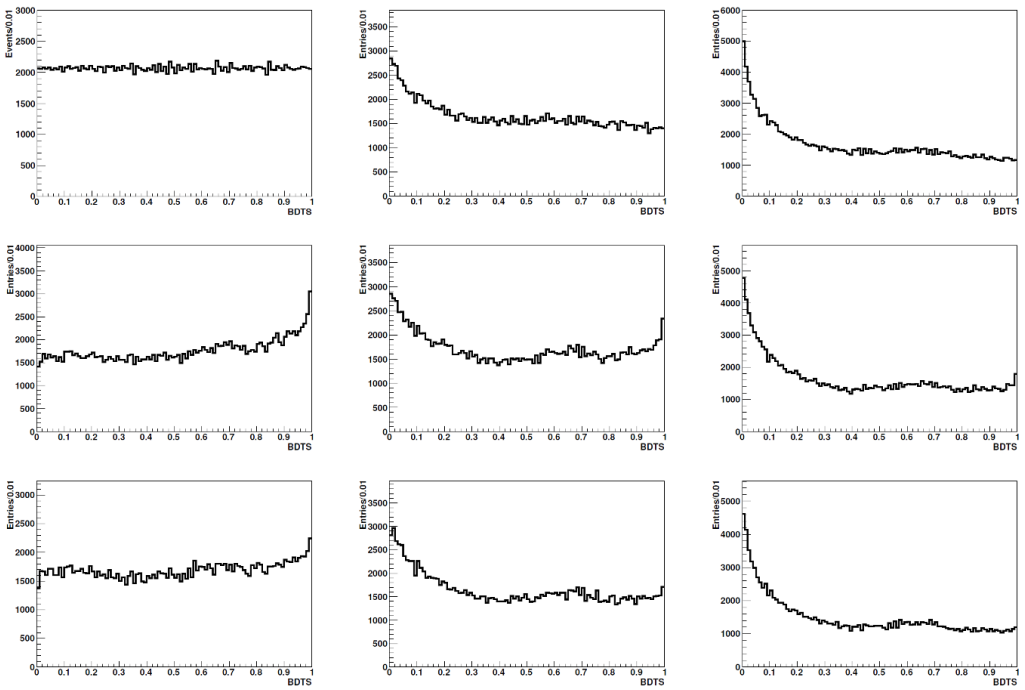

Figure 5.5: BDTS distributions for $B^0_s \to \mu^+\mu^-$ (top), $B^+ \to J/\psi K^+$ (center) and $B^0_s \to J/\psi\phi$ (bottom) using non-smeared (left), smeared (center) and over-smeared (right) Monte Carlo samples.

BDTS cuts are shown in figure 5.7 (left) for $B^0_s \to \mu^+\mu^-$ and $b\bar{b} \to \mu\mu X$ Monte Carlo samples. The curves have been normalised to the number of events before the BDTS cut. Figure 5.7 (right) shows the zoom in the most sensitive region: the blue line obtained with a BDTS cut BDTS>0.05 shows the best performance.

BDTS cut	0.01	0.02	0.03	0.04	0.05	0.08	0.1
$\epsilon_{B_s^0\to\mu^+\mu^-}$	0.9828(3)	0.9664(4)	0.9501(5)	0.9354(6)	0.9210(7)	0.8815(8)	0.8570(9)
$\epsilon_{B^+\to J/\psi K^+}$	0.9834(3)	0.9673(4)	0.9516(5)	0.9372(6)	0.9228(6)	0.8833(8)	0.8587(9)
ratio	1.0006(5)	1.0010(6)	1.0015(8)	1.0019(9)	1.002(1)	1.002(1)	1.002(1)
$\epsilon_{B_s^0\to J/\psi\phi}$	0.9830(3)	0.9650(5)	0.9487(5)	0.9328(6)	0.9170(7)	0.8750(8)	0.8495(9)
ratio	1.0001(5)	0.9986(7)	0.9985(8)	0.9972(9)	0.996(1)	0.993(1)	0.991(1)
$\epsilon_{B^0\to K^+\pi^-}$	0.9825(2)	0.9651(3)	0.9497(4)	0.9350(5)	0.9200(5)	0.8794(6)	0.8538(7)
ratio	0.9996(4)	0.9987(6)	0.9995(7)	0.9995(8)	0.9989(9)	0.998(1)	0.996(1)

Table 5.5: Efficiencies for different BDTS cuts for $B_s^0 \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$ and $B_s^0 \to J/\psi\phi$ smeared Monte Carlo samples. The efficiency ratios between $B^+ \to J/\psi K^+$, $B_s^0 \to J/\psi\phi$, $B^0 \to K^+\pi^-$ and $B_s^0 \to \mu^+\mu^-$ are reported in the third, fifth and seventh row respectively.

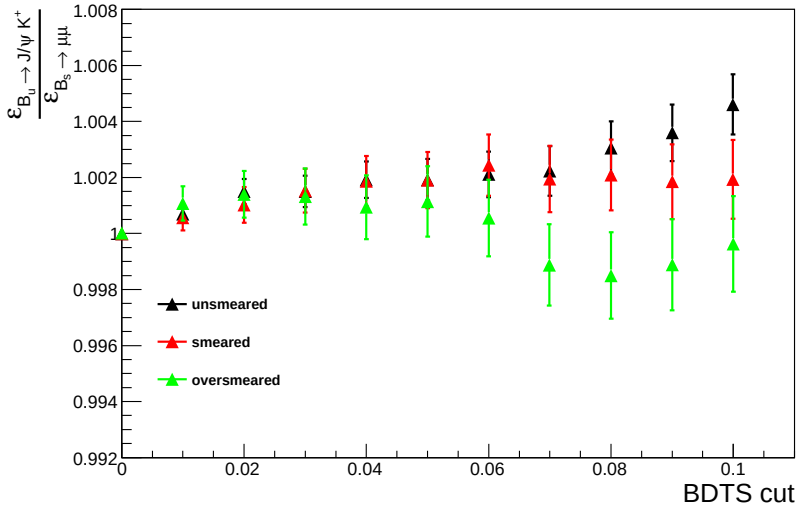

Figure 5.6: Efficiency ratios as a function of the BDTS cut for unsmeared (black triangles), smeared (red triangles) and over-smeared (light green triangles) $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ simulated samples.

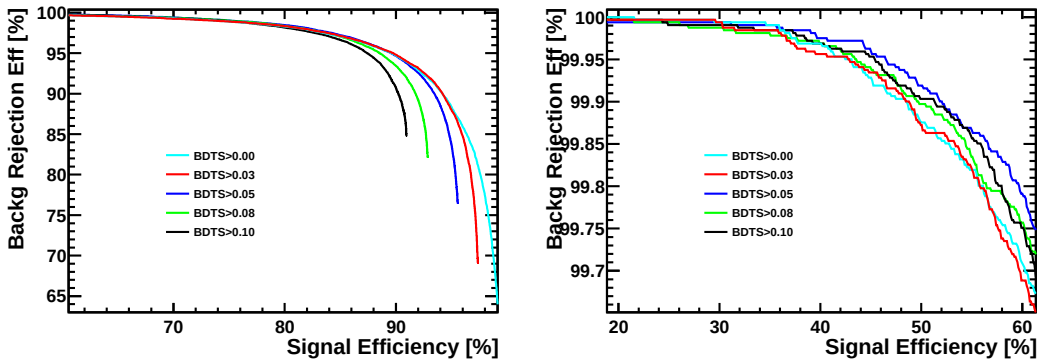

Figure 5.7: BDT Receiver Operating Characteristic (ROC) curves for different BDTS cuts. Left: full range, right: zoomed plot.

5.5 Muon Identification

The muonID procedure is a key ingredient of the $B^0_{(s)} \to \mu^+\mu^-$ analysis. It must be very efficient on $B^0_{(s)} \to \mu^+\mu^-$ signals and has to reduce all the sources of background in which real muons are not present. In addition, the absolute value of the muonID efficiency is required in order to normalise to $B^0 \to K^+\pi^-$ events, while only the ratio of muonID efficiencies is necessary when $B^+ \to J/\psi K^+$ or $B^0_s \to J/\psi\phi$ are used as normalisation channels.

The muonID has been presented in section 3.2.3.3.3, and the references there include wide explanations about the muonID procedure and its performance on 2011 data. A summary of the muonID efficiency results is shown here, focusing in the numbers needed for the $B^0_{(s)} \to \mu^+\mu^-$ analysis.

For the measurement of the muonID efficiency, the tag-and-probe method is applied using $J/\psi \to \mu\mu$ candidates, where one of the muons ($\mu_{\rm tag}$) is identified with IsMuon=1 [124] while the second muon ($\mu_{\rm probe}$) is selected without using any information from the muon system. This second muon is used to estimate the geometrical acceptance of the muon detector relative to tracking chambers, the efficiency of the IsMuon request and the performance of the muon hypothesis test. The muon or non-muon hypothesis is based on how the hits in the muon chambers are aligned with respect to the extrapolation of the tracks from the tracking system. This information is combined together with the informations coming from calorimeters and RICH detectors into a Global Likelihood (DLL) that is used to perform an hypothesis test.

5.5.1 MuonID efficiency

The muonID efficiency is defined here as the number of muons reconstructed as tracks surviving the *IsMuon* requirement with respect to the number of total muons reconstructed as tracks. This efficiency can be factorised into two parts:

1. the first contribution is the geometrical acceptance, α_μ , i.e. the request that the reconstructed muon track points into the muon detector;
2. the second contribution is related to the efficiency of finding hits inside the FOI in the muon stations for tracks pointing into the muon detector, ϵ_μ .

5.5.1.1 Data Samples

The stripping selections used are *JpsiFromBNoPIDNoMip* (selecting $b \to J/\psi X$ events, 2-body sample) and *JpsiKFromBNoPIDNoMip* (selecting $B^+ \to J/\psi K^+$ events, 3-body sample). In order to perform the measurement of the muon detector acceptance, in both lines the request that the $\mu_{\rm probe}$ is in muon acceptance was registered but not enforced. The results obtained with the 2-body sample are then compared with the ones obtained with the 3-body sample and the observed difference added as a systematic uncertainty due to the choice of the sample. This uncertainty has been seen to be almost negligible.

5.5.1.2 Selection

The "tag-and-probe" selection used in muonID efficiency measurement is summarised in Table 5.6.

For the *JpsiKfromB* line only, an extra track with IP$\chi^2 > 25$ is required, and the $\mu\mu K$ invariant mass must be in a 500 MeV/c window around M_{B^+}.

5.5.1.3 Trigger

In order to avoid any bias due to the trigger in the calculation of the muonID efficiency, the $\mu_{\rm probe}$ is required not to fire the trigger at any level ("TIS event with respect to $\mu_{\rm probe}$", "TIS-unbias" in the following)[2]. This concept will be seen later in more detail in section 5.9.5.1. To cross-check the results obtained with this condition, a different sample is used, forcing $\mu_{\rm tag}$ to satisfy the trigger by itself ("TOS event with respect to $\mu_{\rm tag}$" or "TOS-unbias" in the following) [3].

[2] A trigger independent of signal (TIS) event with respect to $\mu_{\rm probe}$ is defined by the following condition: $[L0Global \times Hlt1Global \times Hlt2Global]_{\rm TIS}(\mu_{\rm probe})$ = True.

[3] A trigger on signal (TOS) event with respect to $\mu_{\rm tag}$ is defined by the following specifications: $[L0SingleMuon \times (Hlt1SingleMuonNoIP + Hlt1Track) \times Hlt2SingleMuon]|_{TOS(\mu_{\rm tag})}$=True.

$\mu_{\rm probe}$ and $\mu_{\rm tag}$	two long tracks with opposite charge $\chi^2_{\rm trk}/ndof < 3$ $p > 3$ GeV/c $p_T > 800$ MeV/c IP$\chi^2 > 25$
J/ψ	$\chi^2_{\rm vertex}/ndof < 8$ vertex distance significance > 15 $\lvert M_{\mu\mu} - M_{J/\Psi}\rvert < 200$ MeV/c^2
$\mu_{\rm tag}$ only	$p_T > 1.5$ GeV/c $p > 6$ GeV/c IP > 120 μm MuAcc = 1 IsMuon = 1
$\mu_{\rm probe}$ only	IP > 50 μm MuAcc=1 (ϵ_μ evaluation only)

Table 5.6: Selection cuts for the J/ψ, the $\mu_{\rm tag}$ and the $\mu_{\rm probe}$, used to define the samples for α_μ and ϵ_μ measurements.

The difference between the results obtained with the two samples is added as systematic uncertainty due to the trigger.

5.5.1.4 Method to measure the efficiency and background subtraction

The values of the muonID acceptance, α_μ, and the muonID efficiency given acceptance, ϵ_μ, are extracted from the above samples of unbiased $\mu_{\rm probe}$, properly selecting the events around the J/ψ mass peak. Since the average muonID acceptance or efficiency measured in a given mass window includes contributions from both muons from J/ψ and spurious tracks, a background subtraction procedure has to be applied. Two different methods have been used to evaluate α_μ and ϵ_μ, and two different approaches have been used to subtract the background. This translates into four possible estimates. All the details about the methods are described in [166]. The reference method is the one called $W11$ in [166]. The other methods are used to evaluate the systematic uncertainty on α_μ and ϵ_μ due to the method.

5.5.1.5 Results

Figures 5.8, 5.9 and 5.10 show the acceptance, efficiency and the convolution of the acceptance and efficiency curves as a function of the momentum of $\mu_{\rm probe}$ (left) and the 2D maps (right) as a function of the p and p_T of the probe muon.

In the left plots the results obtained with the 2-body sample are compared with the ones obtained with the 3-body sample and with the corresponding curves from simulated events. The acceptance and efficiency 2D maps are used in the computation of the normalisation factors, as explained in section 5.9. Moreover, these maps will be also used for the normalisation in the $K^0_{\rm S} \to \mu^+\mu^-$ analysis. This will be seen in section 6.7.

5.5.2 Muon hypothesis test (DLL)

As it will be shown in section 5.6.2, the simultaneous application of cuts in the $DLL(K-\pi)$ and $DLL(\mu-\pi)$ variables (section 3.2.3.3) can considerably reduce the number of double misidentified $B^0_{(s)} \to h^+h'^-$ events (i.e, events in which both hadrons are misidentified as muons, faking the $B^0_s \to \mu^+\mu^-$ signal). In particular, the request $DLL(K-\pi) < 10 \& DLL(\mu-\pi) > -5$ can reduce by a factor of $\sim$5 the rate of $B^0_{(s)} \to h^+h'^-$ events with double decays in flight (table 5.7).

The efficiency loss for $B^0_s \to \mu^+\mu^-$ signal has been evaluated by measuring in bins of p of the probe muon from a $J/\psi \to \mu\mu$ sample the efficiency of this set of DLL cuts (ϵ^{data}_{DLLcut}) and then folding it into the spectrum of selected MC $B^0_s \to \mu^+\mu^-$ events. Figure 5.11 (left) shows the efficiency of the DLL

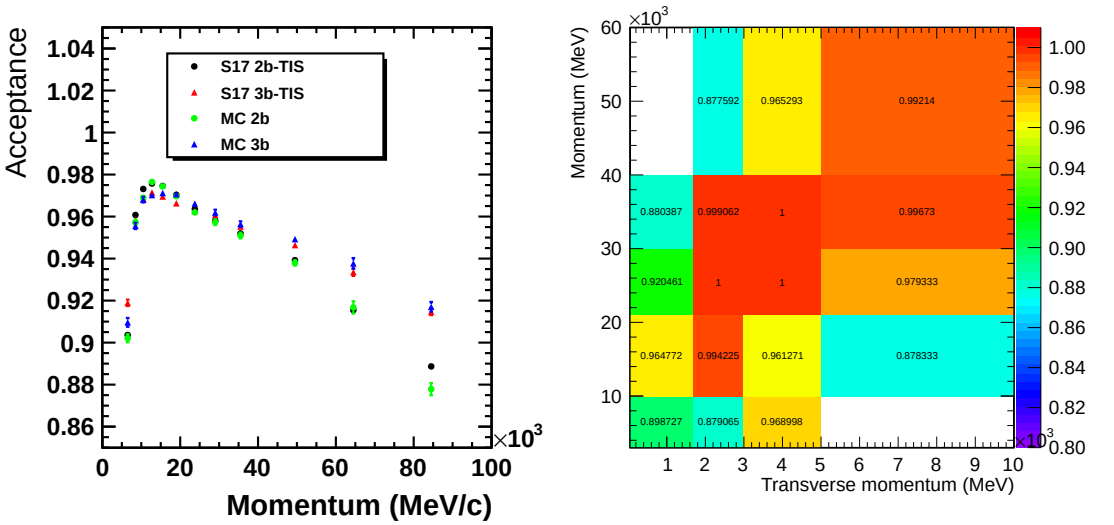

Figure 5.8: Acceptance of the muon system as a function of momentum (p and p_T in the right plot) of the probe muon for 2-body ($b \to J/\psi X$) and 3-body (J/ψ from $B^+ \to J/\psi K^+$) samples. In the left plot, data are compared with Monte Carlo predictions. In the right plot, only the 2-body sample is shown.

cuts $DLL(K-\pi) < 10$ and $DLL(\mu-\pi) > -5$ measured on a sample of $J/\psi \to \mu\mu$ on data and on Monte Carlo as a function of the momentum of μ-probe. The difference between data and Monte Carlo is fully due to the RICH $DLL(K-\pi)$ cut, since the efficiency for $DLL(\mu-\pi) > -5$ agrees with the simulation (figure 5.12).

The efficiency on a $B^0_s \to \mu^+\mu^-$ sample has been evaluated to be 97.3%. The same efficiency has been calculated in bins of the invariant mass-BDT plane, as shown in the right panel of figure 5.11. The efficiency is almost flat as a function of the invariant mass, while a ∼1.5% slope is present when studied as a function of the BDT output; this correction has been applied to the BDT PDF extracted from the $B^0_{(s)} \to h^+h'^-$ sample, as it will be discussed in section 5.8.1.

The tiny effect of the DLL cut on the mass resolution has been studied by fitting the invariant mass of $B^0_s \to \mu^+\mu^-$ simulated sample without and with a DLL cut: when the DLL cut is applied each event is weighted by the efficiency corrections evaluated in data. The resolution of the invariant mass is larger by 0.2 MeV/c^2 for the corrected sample, bias which is well withing the total uncertainty in the mass resolution itself quoted in section 5.8.3.

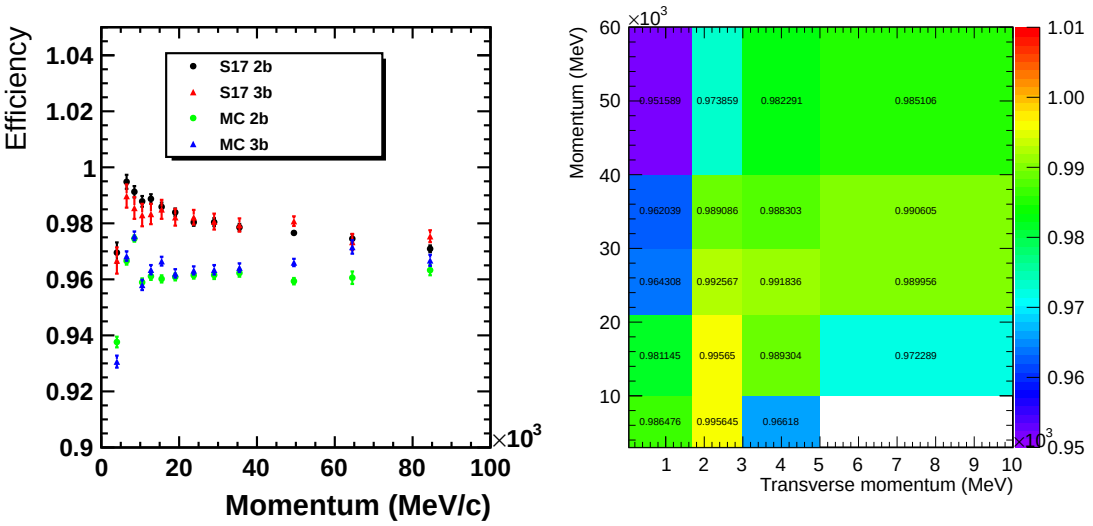

Figure 5.9: Efficiency of the IsMuon=1 requirement as a function of momentum (p and p_T in the right plot) of the probe muon for 2-body ($b \to J/\psi X$) and 3-body (J/ψ from $B^+ \to J/\psi K^+$) samples. In the left plot, data are compared with Monte Carlo predictions. In the right plot, only the 2-body sample is shown.

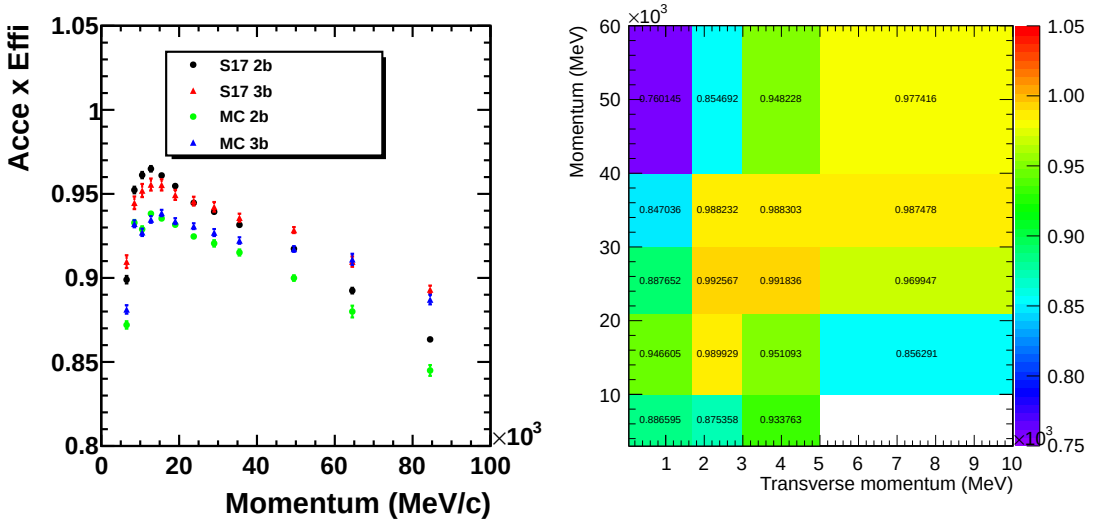

Figure 5.10: Convolution of the acceptance and IsMuon=1 requirements as a function of momentum (p and p_T in the right plot) of the probe muon for 2-body ($b \to J/\psi X$) and 3-body (J/ψ from $B^+ \to J/\psi K^+$) samples. In the left plot, data are compared with Monte Carlo predictions. In the right plot, only the 2-body sample is shown.

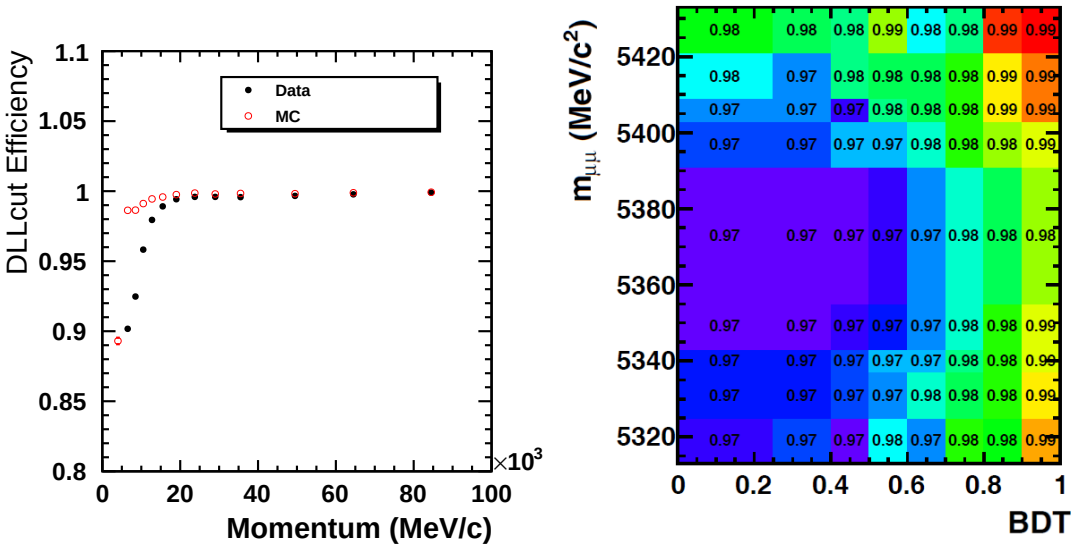

Figure 5.11: Study of the effect of DLL cuts. Left: efficiency of the DLL cut $DLL(K-\pi) < 10 \& DLL(\mu-\pi) > -5$ measured on data and MC samples as a function of the momentum of μ-probe, after the request IsMuon=1. Right: efficiency of the same DLL cut when folded into the p, p_T spectrum of selected $B^0_s \to \mu^+\mu^-$ events as a function of the $B^0_s \to \mu^+\mu^-$ invariant mass and BDT.

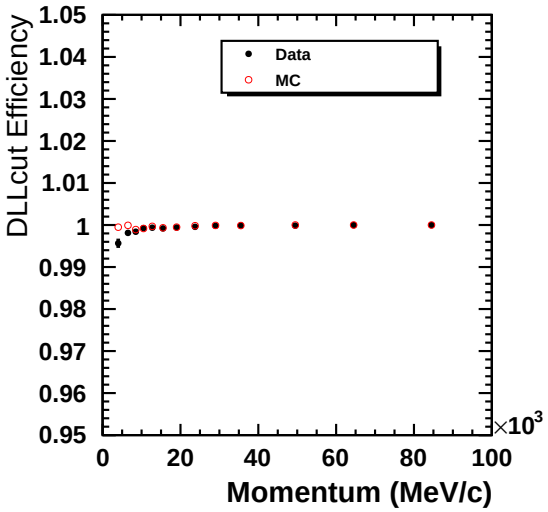

Figure 5.12: Efficiency of the DLL cut $DLL(\mu-\pi) > -5$ measured on data and MC samples as a function of the momentum of μ-probe, after the request IsMuon=1.

5.6 Misidentification rates and background composition

The number of combinatorial background events in the search windows is estimated by interpolating from the mass sidebands, so that the knowledge of the exact background composition is not required. The combinatorial background is fully dominated by the $b\bar{b} \to \mu\mu X$ component: this is shown in figure 5.13, where the $DLL(\mu - \pi)$ distributions[4] and the BDT distributions obtained with background events in data sidebands are compared with the same distributions obtained from simulated $b\bar{b} \to \mu\mu X$ events.

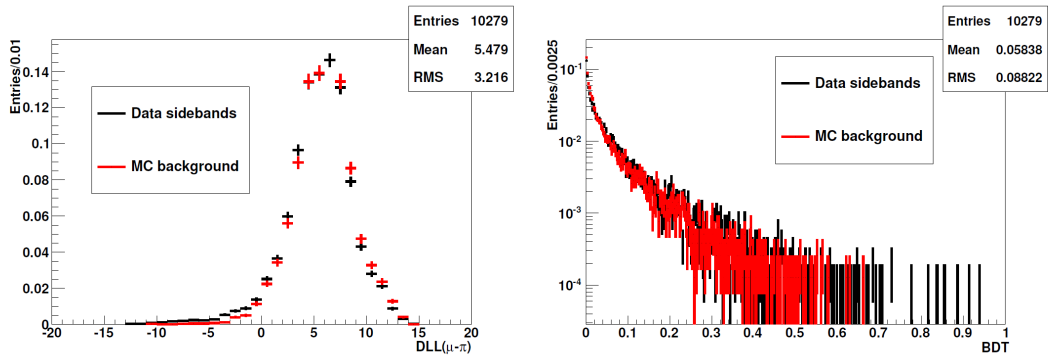

Figure 5.13: $DLL(\mu - \pi)$ and BDT distributions obtained with background events in data sidebands and simulated $b\bar{b} \to \mu\mu X$ events.

Among the peaking backgrounds, the most worrisome is represented by the $B^0_{(s)} \to h^+h'^-$ decays (with $h^{(\prime)}$ being a kaon or a pion) where both hadrons are misidentified as muons since these decays are topologically very similar to the $B^0_{(s)} \to \mu^+\mu^-$ signals.

5.6.1 Pion and kaon misidentification probabilities

The $\epsilon(\pi \to \mu)$ and $\epsilon(K \to \mu)$ fake rates have been measured using a sample of $D^0 \to K\pi$ from the $D^* \to D^0\pi$ decays extracted from ~ 900 pb^{-1} of the stripping selection *NoPIDDstarWithD02RSKPiLine*.

Two different approaches have been used to cancel possible biases arising from the trigger:

- the kaon or pion probe track is required to be TIS with respect the L0Global and HLT1Phys lines ("HLT1TIS" in the following), a similar requirement on HLT2Phys line leaving the result unaffected;
- the kaon or pion probe track is required to be TIS with respect the L0Global, while a *Hlt1AllL0Track* decision is required to the event ("TrackAllDec" in the following), which, in the case of events with hadron decays, gives sufficient unbias with respect to the muon HLT1 lines.

The comparison between the two unbias strategies is used as a systematic cross-check, the first being more accurate from a pure trigger unbias strategy perspective, but also providing less statistics.

In order to cleanup the sample after the stripping selection, the track of the D^0 decays that is not used to evaluate the fake rate is required to be well identified with a cut on the $DLL(K - \pi)$[5]. To evaluate the $\pi \to \mu$ and $K \to \mu$ fake rates two methods have been used:

1. $D^0 \to K\pi$ candidates are selected within a window of $\pm 20\,\mathrm{MeV}/c^2$ around the mass peak and requiring that the π or the K passes the IsMuon=1 condition. The fake rate measured in a given mass window ($\epsilon_{\mathrm{S+B}}$) is related to the actual fake rate ϵ_{S} by the following equation:

$$\epsilon_{\mathrm{S+B}} = \frac{S}{S+B}\epsilon_S + \frac{B}{S+B}\epsilon_B \tag{5.1}$$

[4] DLL refers to the combination of likelihoods from all subdetectors unless otherwise stated.

[5] The kaon (pion) is required to pass the cut $DLL(K - \pi) > 10$ ($DLL(K - \pi) < 0$) when the $\epsilon(\pi \to \mu)$ ($\epsilon(K \to \mu)$) fake rate is measured.

where S and B are the number of signal and background events in a ± 20 MeV/c^2 mass window around the peak, which are extracted by fitting with a Crystal Ball the signal and with a polynomial function the background. The fake rate for the background $\epsilon_{\rm B}$ is evaluated in the mass sidebands ($|m - m_{D^0}| > 40$ MeV/c^2). The above relation becomes:

$$\epsilon_{\rm S} = \frac{B+S}{S}\left(\epsilon_{S+B} - \frac{B}{S+B}\epsilon_B\right) \tag{5.2}$$

solving for $\epsilon_{\rm S}$. The procedure is repeated by dividing the D^0 sample in 11 bins in p and 4 bins in p_T of the track used to evaluate the fake rates (probe track). In figure 5.14 the mass distributions corresponding to 10 p-bins for a p_T range $1700 < p_T < 3000$ MeV/c of the probe track are shown.

2. The mass distributions of D^0 candidates are obtained by dividing the sample in 11 bins in p and 4 bins in p_T of probe track as in the previous method. After requiring the probe track to satisfy the condition IsMuon = 0, each distribution is fitted and the number of signal candidates N_S(IsMuon=0) obtained. The fits are then repeated by requiring the probe track to satisfy IsMuon=1 and N_S(IsMuon=1) is extracted. Each distribution is fitted with a double Gaussian for the signal and an exponential function for the background (an alternative fit model with a single Gaussian for the signal and a 3rd order Chebyshev polynomial for the background has been also used). Finally the fake rate is obtained via the relation:

$$\epsilon_S = \frac{N_S(\text{IsMuon} = 1)}{N_S(\text{IsMuon} = 0) + N_S(\text{IsMuon} = 1)} \tag{5.3}$$

The fits for 10 bins in p of the probe track for the range $1700 < p_T < 3000$ MeV/c are shown in figure 5.15 for IsMuon=0 and in figure 5.16 for IsMuon=1.

This method has been also implemented, for systematic evaluation, by changing the strategy for signal counting (in both IsMuon=0 and IsMuon=1 hypotheses) to a background extrapolation under the peak via an exponential function.

The fake kaon misID probability evaluated with method 1 in HLT1TIS hypothesis as a function of probe track momentum for the four p_T bins is shown in figure 5.17 (top); the same probability computed with method 2 in TrackAllDec hypothesis is shown in the middle, and the ratio between the two in the bottom. The same results are shown for pion in figure 5.18. The two methods are in good agreement. A systematic error has been assessed by using both set of curves to compute the double misID background.

In order to reduce the number of double misidentified $B^0_{(s)} \to h^+h'^-$ events, cuts on the $DLL(K-\pi)$ and $DLL(\mu-\pi)$ can be applied on top of the IsMuon requirement. In section 5.5 it was demonstrated that the combination of cuts $DLL(K-\pi) < 10$ and $DLL(\mu-\pi) > -5$ selects $B^0_s \to \mu^+\mu^-$ with high efficiency, $\sim$0.97. The same selection, applied to kaons and pions, reduces the misID probabilities by large factors, as shown in figure 5.19.

5.6.2 Double misidentification probability

The double fake rate has been estimated by convoluting the kaon and pion misID curves given above with the momentum and p_T spectrum of the two hadrons of MC $B^0_{(s)} \to h^+h'^-$ decays, selected as described in section 5.4. This has been done with a toy technique that takes properly into account the uncertainties on the fake rate in each of the 4×11 bins and the correlations between p and p_T of both tracks. The average double misID probability for the exclusive $B^0_{(s)} \to h^+h'^-$ decays is shown in table 5.7 before and after the DLL selection described above, and separately for the HLT1TIS and TrackAllDec samples.

The average double misID for inclusive $B^0_{(s)} \to h^+h'^-$ decays, $\epsilon_{hh\to\mu\mu}$, is also shown in table 5.7, and it has been obtained by weighting the values obtained for exclusive decays according to their relative production rate [6].

Before DLL cuts, a double misID rate $\epsilon_{hh\to\mu\mu} = (8.6 \pm 0.34(\text{stat}) \pm 0.30\ (\text{syst})) \times 10^{-5}$ is measured, where the systematic error comes from the difference between the results given in table 5.7, plus an additional contribution coming from fit models. After DLL cuts, the double

[6] The following branching fractions are assumed: $\mathcal{B}(B^0 \to K^+\pi^-) = (1.95 \pm 0.06) \times 10^{-5}$, $\mathcal{B}(B^0 \to \pi^+\pi^-) = (5.13 \pm 0.24) \times 10^{-6}$, $\mathcal{B}(B^0_s \to \pi^+K^-) = (5.05 \pm 1.0) \times 10^{-6}$, $\mathcal{B}(B^0_s \to K^+K^-) = (3.5 \pm 0.7) \times 10^{-5}$ and $f_s/f_d = 0.267 \pm 0.021$.

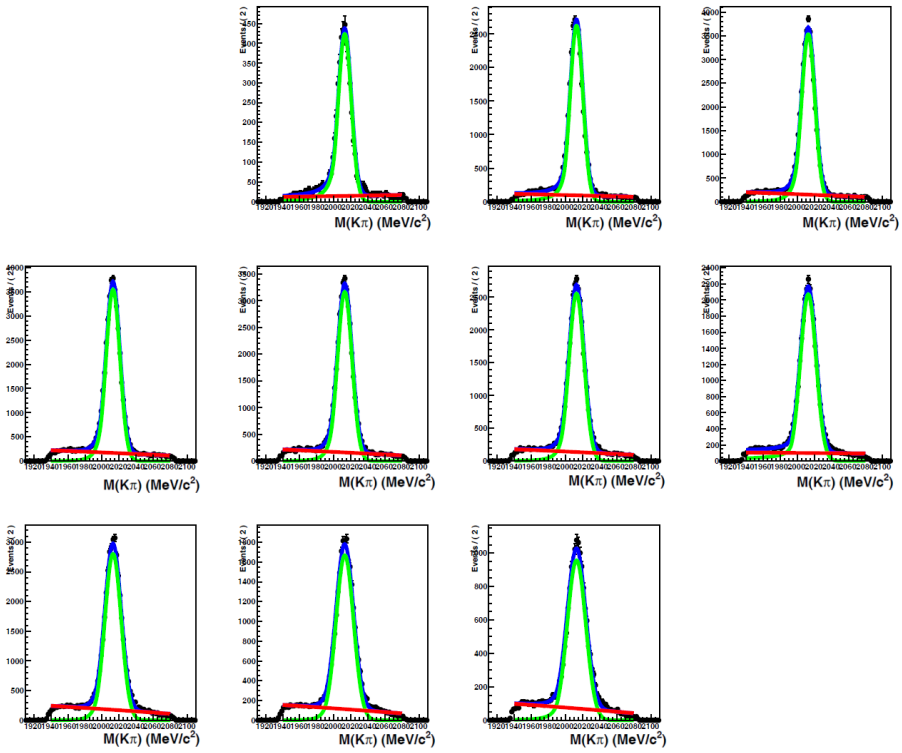

Figure 5.14: Mass distributions of the $D^0 \to K\pi$ decay for 10 p bins and for the p_T range 1700 MeV/c $< p_T <$ 3000 MeV/c of the probe track used to estimate the fake rate $\epsilon(K \to \mu)$ with method 1.

misID rate is reduced by more than a factor of 5, giving $\epsilon_{hh\to\mu\mu} = (1.52 \pm 0.07_{\text{stat}} \pm 0.07_{\text{syst}}) \times 10^{-5}$.

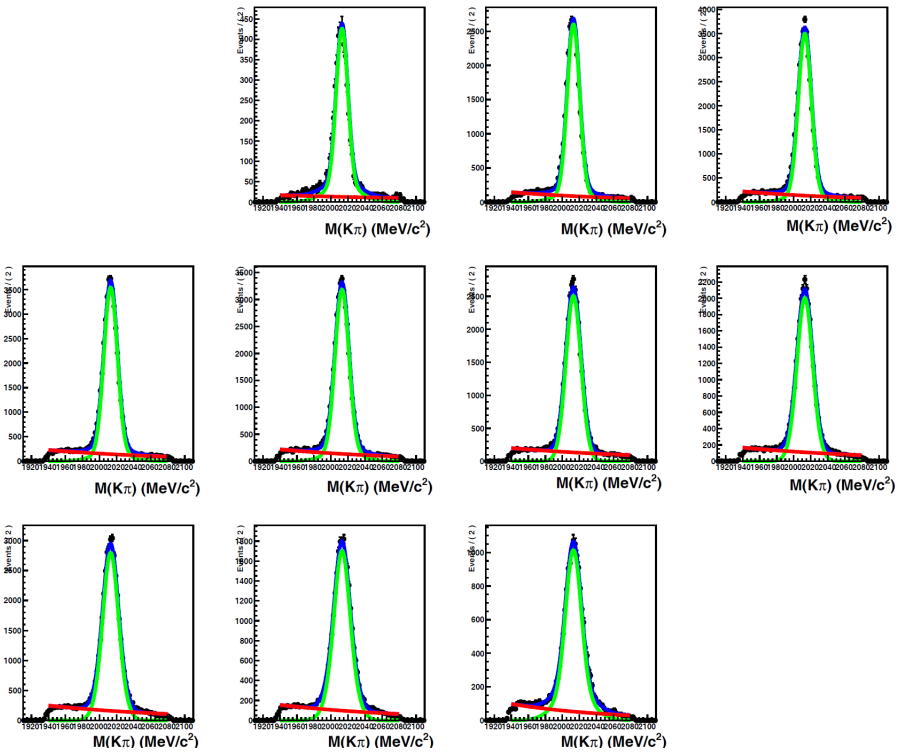

Figure 5.15: Mass distributions of the $D^0 \to K\pi$ decay for 10 p bins and for the p_T range 1700 MeV/c $< p_T <$ 3000 MeV/c of the probe track used to estimate the fake rate $\epsilon(K \to \mu)$ with method 2. The probe track is required to satisfy the condition IsMuon=0.

dataset	$B^0_s \to K^+K^-$ $\times 10^{-4}$	$B^0_s \to \pi^+K^-$ $\times 10^{-4}$	$B^0 \to K^+\pi^-$ $\times 10^{-4}$	$B^0 \to \pi^+\pi^-$ $\times 10^{-4}$	$B^0_{(s)} \to h^+h'^-$ $\times 10^{-4}$
HLT1TIS	1.19 ± 0.05	0.82 ± 0.05	0.84 ± 0.05	0.52 ± 0.03	0.88 ± 0.04
TrackAllDec	1.16 ± 0.04	0.79 ± 0.05	0.81 ± 0.04	0.50 ± 0.03	0.86 ± 0.03
$DLL(K-\pi) < 10$ and $DLL(\mu-\pi) > -5$					
HLT1TIS	0.065 ± 0.006	0.146 ± 0.008	0.146 ± 0.008	0.334 ± 0.014	0.153 ± 0.008
trackAllDec	0.066 ± 0.005	0.148 ± 0.007	0.148 ± 0.007	0.323 ± 0.012	0.152 ± 0.007

Table 5.7: Average double misID probability for exclusive $B^0_{(s)} \to h^+h'^-$ decays and for the inclusive combination of them, for the two series of single particle misID probabilities available.

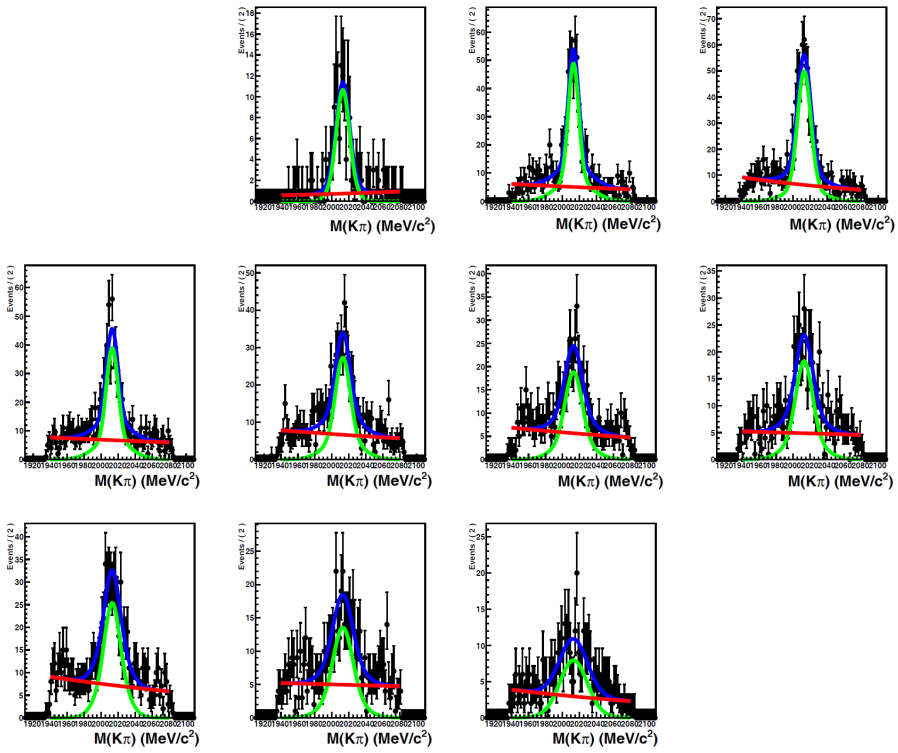

Figure 5.16: Mass distributions of the $D^0 \to K\pi$ decay for 10 p bins and for the p_T range 1700 MeV/c $< p_T <$ 3000 MeV/c of the probe track used to estimate the fake rate $\epsilon(K \to \mu)$ with method 2. The probe track is required to satisfy the condition IsMuon=1.

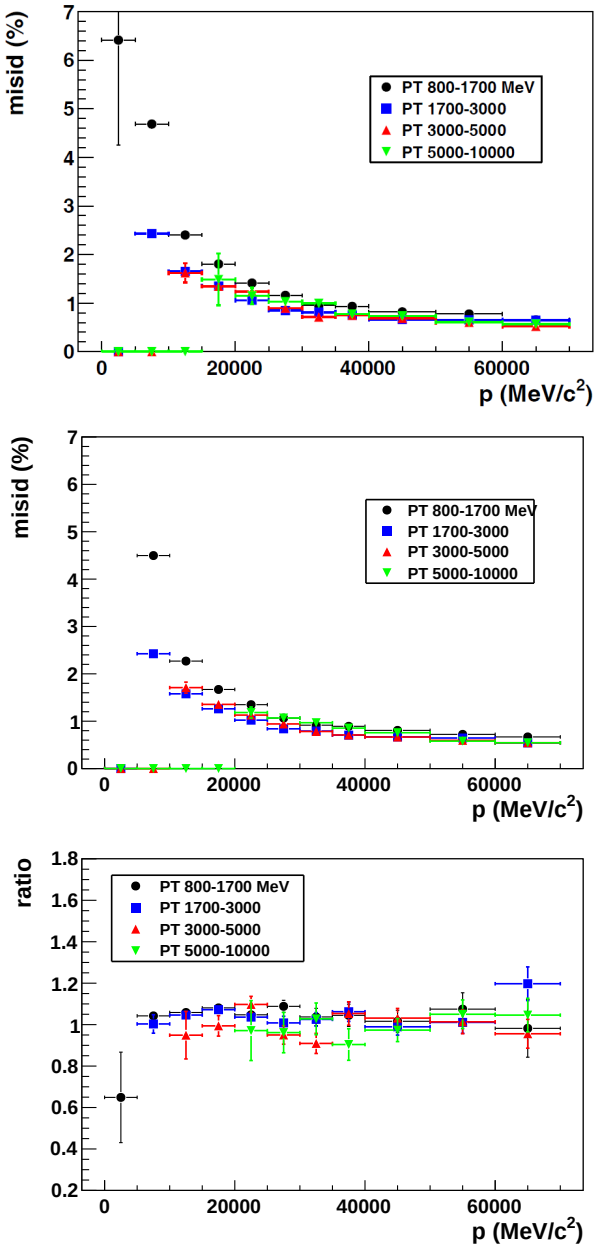

Figure 5.17: Kaon misID vs. momentum for the four p_T bins of the probe track: HLT1TIS sample (top), TrackAllDec sample (middle), and the ratio between the two.

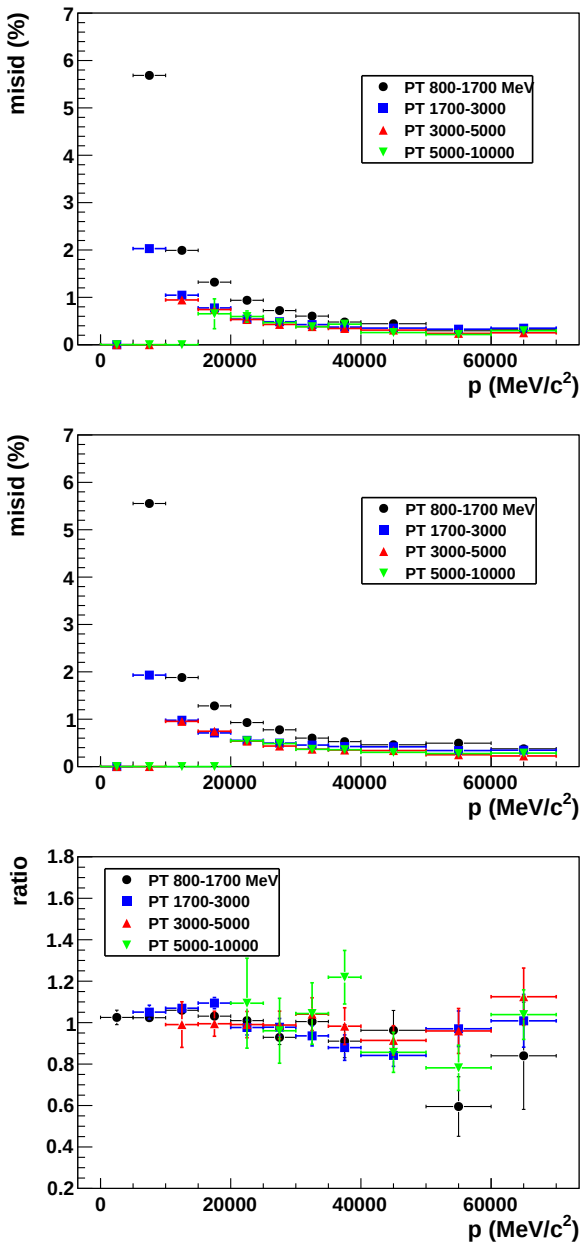

Figure 5.18: Pion misID vs. momentum for the four p_T bins of the probe track: HLT1TIS sample (top), TrackAllDec sample (middle), and the ratio between the two.

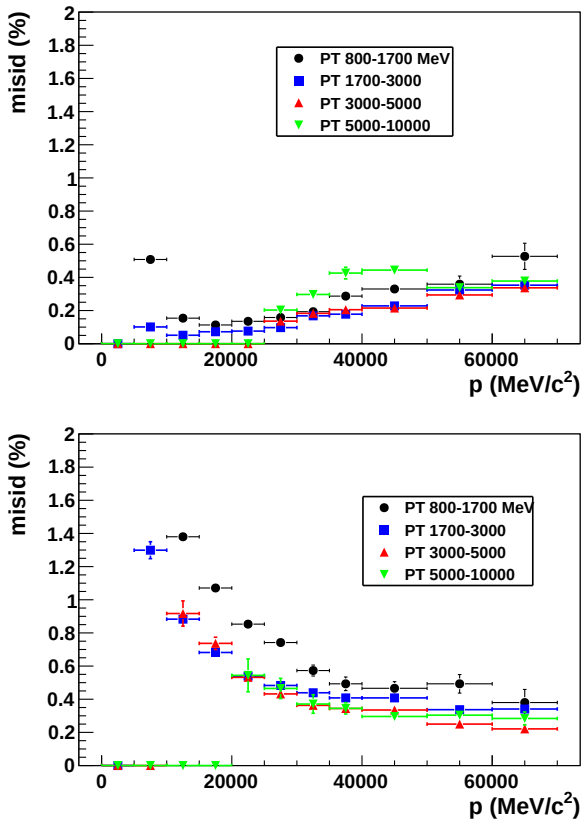

Figure 5.19: Kaon (top) and pion (bottom) misID probabilities vs. momentum for the four p_T bins of the probe track after $DLL(K-\pi) < 10 \& DLL(\mu-\pi) > -5$ cuts (obtained with method 2 and TrackAllDec trigger selection)

5.6.3 Peaking background from $B^0_{(s)} \to h^+h'^-$ with double misidentification: yield, BDT and mass PDFs

The number of $B^0_{(s)} \to h^+h'^-$ double misidentified events is evaluated as:

$$N^{\rm TRIG|SEL}_{B^0_{(s)}\to h^+h'^-\to\mu\mu} = \epsilon^{\rm TRIG|SEL}_{B^0_{(s)}\to\mu^+\mu^-}\ \frac{N^{\rm TIS}_{hh}}{\epsilon^{TIS}\epsilon^{HLT2,MC}}\ \epsilon_{hh\to\mu\mu} \tag{5.4}$$

where $N^{\rm TIS}_{hh}$ is the number of $B^0_{(s)} \to h^+h'^-$ TIS events[7], ϵ^{TIS} is the L0 and HLT1 TIS efficiency, $\epsilon^{HLT2,MC}$ is the HLT2 TOS efficiency and $\epsilon^{\rm TRIG|SEL}_{B^0_s\to\mu^+\mu^-}$ is the $B^0_{(s)} \to \mu^+\mu^-$ trigger efficiency. These efficiencies are estimated in section 5.9.

As a result, the expected background events due to double misID of $B^0_{(s)} \to h^+h'^-$ events before DLL cuts is (30.4 ± 4.2) in the whole BDT range, for an integrated luminosity of 1.02 fb^{-1}. After the DLL cuts the double misID events are estimated to be (5.4 ± 0.7).

A second estimation of the peaking background is obtained from data as a cross-check. The average single misID rate per hadron has been computed in a sample of $B^0_{(s)} \to h^+h'^-$ L0TIS from events passing Hlt1AllL0Track, in which one of the hadrons has been misidentified as muon, later squaring that value to obtain the double misID rate. Events are selected with BDT > 0.5 and muon candidates are required to pass the DLL cuts. The estimated single muonID fake rate per hadron ($\epsilon_{h\to\mu}$) is obtained by inverting the formula:

$$p_{h\mu/\mu h} = 2 \times \epsilon_{h\to\mu} \times (1 - \epsilon_{h\to\mu}) \sim 2 \times \epsilon_{h\to\mu} \tag{5.5}$$

where $p_{h\mu/\mu h}$ is the probability to have at least a hadron misidentified as muon:

$$p_{h\mu/\mu h} = \frac{N_{h\mu/\mu h}}{N_{h\mu/\mu h} + N_{hh}} \tag{5.6}$$

and where $N_{h\mu/\mu h}$ is the number of $B^0_{(s)} \to h^+h'^-$ events with a hadron misidentified as muon and N_{hh} is the number of $B^0_{(s)} \to h^+h'^-$ events with no hadron misidentified as muon. Both N_{hh} and $N_{\mu h/h\mu}$ are extracted from a fit of the invariant mass distribution of L0TIS $B^0_{(s)} \to h^+h'^-$ and which come from events passing Hlt1AllL0Track, as explained before. The $B^0_{(s)} \to h^+h'^-$ candidates are separated by requiring none of the hadrons to have been misidentified (figure 5.20, right) or only one of the hadrons to have been misidentified (figure 5.20, left). Two methods to extract the number of $B^0_{(s)} \to h^+h'^-$ events are used: the first of 1D fits used in the BDT calibration, explained in section 5.8.1 (figure 5.20, top) and computing the events above the background, modeled as an exponential (figure 5.20, bottom); both methods give compatible numbers.

The single hadron misID rate for DLL cuts (with $BDT > 0.5$) is: $\epsilon_{h\to\mu} = (0.33 \pm 0.05_{stat} \pm 0.10_{syst}) \times 10^{-2}$. The systematic error is due to the estimation of the background contamination under the mass peak. The stability of the misID (with DLL cuts) rate vs. time has been checked in two different run periods. The double misID rate (with DLL cuts) is $(1.11 \pm 0.31\ (\text{stat}) \pm 0.57\ (\text{syst})) \times 10^{-5}$, which is in agreement with the value of misID rate obtained with the method described above in the same range of BDT, $(1.44 \pm 0.09) \times 10^{-5}$.

To take into account the dependence of double misID from the BDT bins, this has been evaluated as a function of the BDT bin, $\epsilon_{hh\to\mu\mu}(i)$, and a fractional correction for each BDT bin has been defined as $f_{misID}(i) = \epsilon_{hh\to\mu\mu}(i)/\epsilon_{hh\to\mu\mu}$, where $\epsilon_{hh\to\mu\mu}$ is the average double misID given above. The values of $f_{misID}(i)$ are given in table 5.8, before and after the DLL cuts; the values are given without errors, since they are mostly correlated among the bins. This correction is then introduced in equation 5.4, as well as the dependence of the trigger factors (ϵ^{TIS}, $\epsilon^{HLT2,MC}$) on the BDT bin (also discussed in section 5.8.1).

[7] The inclusive $B^0_{(s)} \to h^+h'^-$ TIS events are evaluated with tracks in the geometrical acceptance of the muon detector, matching the definition of the above misID probabilities.

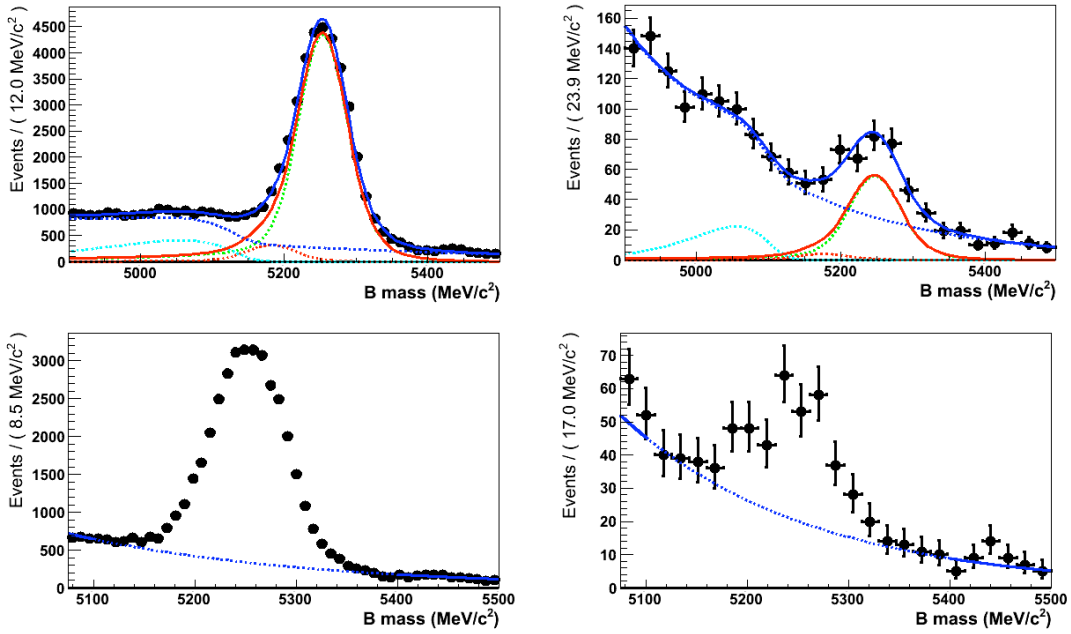

Figure 5.20: Estimation of double misID in data. Left, invariant mass distribution of $B^0_{(s)} \to h^+h'^-$ events in which none of the hadrons has been identified as muon; right, invariant mass distribution of $B^0_{(s)} \to h^+h'^-$ events in which only one of the hadrons has been identified as muon (i.e., fullfills IsMuon and the DLL cuts). Top, results of the full fit (first of 1D fits used for the estimation of the BDT PDF of the signal, section 5.8.1); botton, fit to the background only (fit to an exponential). In all cases, B candidates are L0TIS and come from events passing Hlt1AllL0Track.

To estimate the number of the peaking background events that are in the B^0 and B^0_s signal mass window, the invariant mass lineshape of $B^0_{(s)} \to h^+h'^-$ with a double decay in flight has also been evaluated. In this way, the change of mass hypothesis from $h^+h'^-$ to $\mu^+\mu^-$, together with the degradation in the momentum resolution of the misidentified hadron (due to the presence of decays in flight) may move a fraction of this background out of the B^0_s and B^0_s mass windows. The slope of the tracks at the origin (which defines the angle between the tracks) is assumed to be unaffected by the decay in flight while the momentum used for the invariant mass computation is corrected following Monte Carlo simulation. The difference on the momentum of the hadron before and after the decay in flight is obtained by MC and applied to data. The invariant mass distribution for $B^0_{(s)} \to h^+h'^-$ with double decay in flight is obtained by applying the MC momentum corrections to both hadrons, and is shown in figure 5.21.

With this method, the fraction of double misID $B^0_{(s)} \to h^+h'^-$ that are in the search windows

BDT bin	0-0.25	0.25-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9 1.0
$f_{\text{misID}}(i)$	1.17	1.15	1.12	1.07	0.99	0.86	0.74	0.66
				$DLL(K-\pi) < 10$ and $DLL(\mu-\pi) > -5$				
$f_{\text{misID}}(i)$	1.04	1.07	1.06	1.07	1.04	0.95	0.87	0.81

Table 5.8: Ratio between double misID computed in each BDT bin and its average value, before and after the DLL cuts.

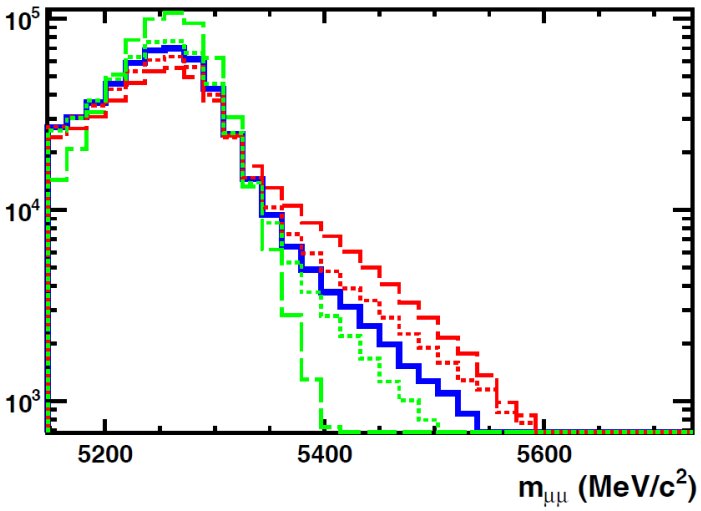

Figure 5.21: Invariant mass distribution calculated for $B^0_{(s)} \to h^+h'^-$ with double misID, evaluated as described in the text: blue curve represents the central value, red and green curves the upper and lower limits. Short dashed curves show how the central curves are modified by moving the track parameters by $\pm 1\,\sigma$ with respect to the central values. Long dashed curves represent the most conservative limits: no kink at all, just wrong mass hypothesis and mass resolution taken from data (green curve) and kink as big as for kaons applied also to pions (red curve).

is found to be $(8.8^{+3.0}_{-2.1})\%$ in ± 60 MeV/c^2 around the B_s mass and $(48.0^{+20}_{-8})\%$ in ± 60 MeV/c^2 around the B_d mass.

5.6.4 Other exclusive backgrounds

5.6.4.1 Background from $B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$

Among the various exclusive backgrounds in the search for $B^0_{(s)} \to \mu^+\mu^-$, the $B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$ appears to be one of the most dangerous. The muon from the B_c^+ semileptonic decay and the oppositely charged muon from the J/ψ can be combined together to form a fake $B^0_{(s)} \to \mu^+\mu^-$ decay. While all the other B semileptonic decays, due to the missing particles when the $B^0_{(s)} \to \mu^+\mu^-$ candidate is reconstructed, fall outside the signal mass region, the large B_c^+ mass can lead the fake signal to lie in it. Moreover, the two muons come exactly from the same decay vertex so that the combination cannot be discarded by the vertex fit or the DOCA between the tracks. Furthermore, this decay could influence the shape of the left sideband leading to an incorrect estimate of the combinatorial background. For these reasons a study has been done with a Monte Carlo sample of this decay.

The different signal selection steps have been applied to this sample. The statistics obtained in the various steps is reported in table 5.9.

The number of expected fake $B_s^0 \to \mu^+\mu^-$ events coming from $B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$ is calculated as follows:

$$N^{exp} = \mathcal{L} \cdot \sigma(pp \to b\bar{b}) \cdot 2 \cdot f_c \cdot \mathcal{B}(B_c^+ \to J/\psi\mu^+\nu) \cdot \mathcal{B}(J/\psi \to \mu^+\mu^-) \cdot \varepsilon \qquad (5.7)$$

where the efficiency includes reconstruction and selection (from table 5.9) and the acceptance efficiency, which has been measured from the Monte Carlo generator to be $\varepsilon_{acc} = 0.1389 \pm 0.0003$. The considered decay branching ratio multiplied by the fragmentation fraction $b \to B_c$ has been

	N	ε
Tot Events	1007188	
After stripping	4315	0.428%
of which MC associated	4216	0.419%
BDTS and fiducial cuts	1917	0.190%
of which MC associated	1900	0.188%
in B^0_s mass window (5358 ± 60) MeV	10	$9.43 \cdot 10^{-6}$
in B^0 mass window (5272 ± 60) MeV	21	$2.085 \cdot 10^{-5}$
in sidebands range [4900 - 6000] MeV	263	$2.611 \cdot 10^{-4}$

Table 5.9: Event statistics obtained on the $B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$ Monte Carlo sample. Listed efficiencies include trigger and stripping, while the acceptance efficiency is already included in the initial sample.

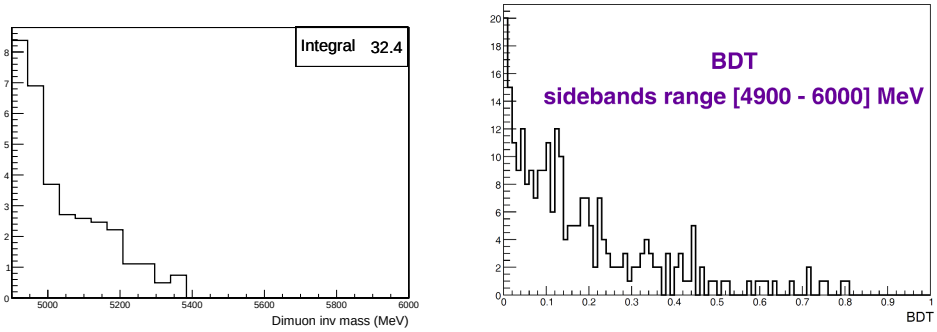

Figure 5.22: $B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$ background. Left, invariant mass distribution of fake $B^0_{d,s} \to \mu^+\mu^-$ candidates from $B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$ scaled to the expected statistics as obtained in Monte Carlo simulations. Right, BDT distribution obtained from the same events (but not scaled). See text for a detailed explanation of these plots.

measured by the CDF collaboration in an inclusive way and considering both a muon and an electron in the final state [167]; the measured value is:

$$f_c \cdot \mathcal{B}(B_c^+ \to J/\psi \ell^+ \nu X) = 5.2^{+2.4}_{-2.1} \cdot 10^{-5}$$

in the absence of more information, the criterion from the authors of this measurement is followed, applying a multiplicative factor of $1/2$ for the muon channel, and considering negligible the non exclusive decays, all this leading to:

$$f_c \cdot \mathcal{B}(B_c^+ \to J/\psi \mu^+ \nu) = 0.5 \cdot 5.2^{+2.4}_{-2.1} \cdot 10^{-5} \quad .$$

Considering the cross-section $\sigma(pp \to b\bar{b}) = (284 \pm 20 \pm 49)$ μb measured by LHCb [150] and the branching ratio $\mathcal{B}(J/\psi \to \mu^+\mu^-) = (5.93 \pm 0.06) \cdot 10^{-2}$ [5], a total expected number of background events of 234 ± 117 in the full stripping mass region has been computed; this number goes down to 32 ± 16 inside the mass region $[4900, 6000]$ MeV. Inside the B^0_s and B^0_d mass windows 1.23 ± 0.73 and 2.59 ± 1.41 events, respectively, are estimated. Note that these estimates are ultimately limited by the MC statistics (which corresponds approximately to about 10 fb^{-1}) but even more by the error on B_c^+ production and branching ratio. The invariant mass distribution of the fake $B^0_{d,s} \to \mu^+\mu^-$ candidates from $B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$, scaled at the expected value, is shown in figure 5.22 left.

The distribution of the BDT variable of the selected events in the invariant mass region

[4900, 6000] is shown, not normalised, in figure 5.22 right. There, it can be seen this distribution peaks at low values of BDT so that the expected events in the B_s^0 and B^0 mass windows are likely to lie in the low BDT region. Taking a BDT value larger than 0.8, only one MC event is left in the B^0 signal region, leading to an expected number of events of 0.12 ± 0.13, while no events are left in the B_s^0 mass window, leading to negligible expected number of events.

Therefore the $B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$ decay is considered as negligible in the signal region, for the present statistics. Finally, its invariant mass distribution in the sidebands is exponentially distributed, so that this decay is largely included in the combinatorial background extrapolation.

5.6.4.2 Background from $B^+ \to \pi^+\mu^+\mu^-$

As seen in section 4.1.3, $B^+ \to \pi^+\mu^+\mu^-$ is a very rare decay recently discovered by LHCb, with a branching fraction $\mathcal{B}(B^+ \to \pi^+\mu^+\mu^-) = (2.4 \pm 0.6 \pm 0.2) \cdot 10^{-8}$. The two muons from this decay could be a background to the $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ decays. While, due to the missing pion, the $\mu\mu$ invariant mass will not be in the signal windows, it could influence the evaluation of the combinatorial background.

The influence of the $B^+ \to \pi^+\mu^+\mu^-$ to this analysis was studied with MC simulations (MC11 production). The total number of expected events as background for $B_s^0 \to \mu^+\mu^-$ is:

$$N^{bkg}_{B^+\to\pi^+\mu^+\mu^-} = \mathcal{L} \cdot 2 \cdot \sigma(pp \to b\bar{b}) \cdot f_u \cdot \mathcal{B}(B^+ \to \pi^+\mu^+\mu^-) \cdot \varepsilon \tag{5.8}$$

Assuming the already mentioned luminosity and $b\bar{b}$ production cross-section, and considering the B^+ production fraction from [5], $f_u = (40.3 \pm 1.1)\%$, the total number of produced N_{B^+} is:

$$N_{B^+} = 2.335 \cdot 10^{11}$$

The efficiencies for $B^+ \to \pi^+\mu^+\mu^-$ as obtained from MC are reported in table 5.10. In summary, after all the cuts (trigger, stripping, fiducial cuts, BDTS and μ-ID)

$$N^{bkg}_{B^+} = N_{B^+} \cdot \mathcal{B}(B^+ \to \pi^+\mu^+\mu^-) \cdot \varepsilon = 12.7$$

events are expected due to $B^+ \to \pi^+\mu^+\mu^-$ in the full invariant mass region and full BDT range.

	N	ε	ε_{rel}
Tot	766200		
Stripping	19157	0.0250026	
Trigger	17847	0.0232929	0.928612
Fiducial cuts	17425	0.0227421	0.976355
BDTS	11506	0.015017	0.660316
DLL cuts	11392	0.0148682	0.990092
BDT>0.5	2583	0.00337118	0.226738

Table 5.10: Efficiencies for the $B^+ \to \pi^+\mu^+\mu^-$ selected as $B_s^0 \to \mu^+\mu^-$. Note: the stripping numbers include also acceptance and reconstruction.

The BDT distribution for the selected $B^+ \to \pi^+\mu^+\mu^-$ candidates is shown in figure 5.23 left, while the mass distribution of the selected candidates, in the full BDT range and scaled to the number of expected candidates, can be found in figure 5.23 right.

In the most sensitive BDT region ($BDT > 0.5$) a total of 2.87 $B^+ \to \pi^+\mu^+\mu^-$ events is expected, among which 0.229 are in the left sideband region $[4900, 5000] MeV/c^2$. No MC events are found in the signal invariant mass regions.

As a final note, the isospin-conjugate mode $B^0 \to \pi^0\mu^+\mu^-$ has also to be considered, since the branching fraction should be approximately the same of the charged one and so should

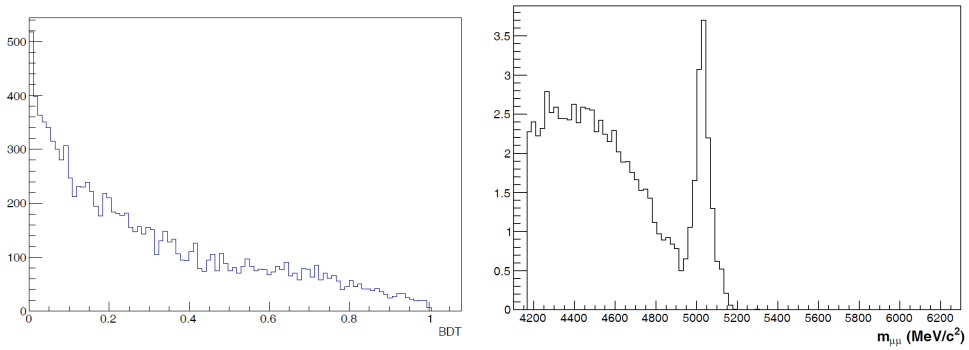

Figure 5.23: The $B^+ \to \pi^+\mu^+\mu^-$ background. Left, BDT distribution. Right, mass spectra of selected candidates in the full BDT range and normalised to the expected yields in the analysis dataset.

be the efficiencies. Therefore, a factor of two for the numbers listed in this paragraph can be roughly assumed, so that this background is still clearly negligible for the present statistics.

5.6.4.3 Background from $B_s \to \mu^+\mu^-\gamma$

A possible background contribution from $B_s \to \mu^+\mu^-\gamma$ has also been considered for this analysis. In this case, the categorisation of this decay as "background" depends on the origin of the photon at Feynman diagram level, as seen in section 2.1.3.3. In this way, three different contributions could be accounted for: photons from initial state radiation (ISR), photons from final state radiation (FSR) and photons coming from an interference between FSR and ISR. ISR and the interference between ISR and FSR can be considered as a background for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, while FSR is essentially part of the signal being looked for.

It is also remarkable that the importance of this background will be very dependent on the branching ratio of $B_s \to \mu^+\mu^-\gamma$ and on the mass range in which $B_s \to \mu^+\mu^-\gamma$ lies. For the former, the branching ratio is not known, but as mentioned in section 2.1.3.3 it is currently thought to be ∼10 times larger than that of $B^0_s \to \mu^+\mu^-$. For the latter, it will vary with the momentum spectrum of the photon, which is difficult to parameterise, and also depending on the origin of the photon (ISR, FSR or interference). For this study, a new parameterisation for the photons for the MC11 simulation has been used, following what it was seen in section 2.1.3.3. It must be said that this MC11 simulation includes, at the same time, all the three mentioned contributions for the photon origin, so that all the results shown are considered to be upper limits for this exclusive background computation, since the contribution from FSR should not be taken into account.

In order to estimate the contribution of this background, two samples of MC11 simulation have been employed. For $B_s \to \mu^+\mu^-\gamma$, as just explained, a set of specially generated 1.66 M events has been used. At the same time, 0.21 M $B^0_s \to \mu^+\mu^-$ events have also been studied. With this, the reconstruction, selection and trigger efficiencies have been obtained purely on MC. For the selection, the cuts explained in section 5.4 have been applied. For the trigger, L0, HLT1Physics[8] and HLT2Physics decisions were required. On top of this, the focus has been put in last two BDT bins (BDT>0.8, as it will be seen in section 5.7.2). The mass spectrum of the simulated $B_s \to \mu^+\mu^-\gamma$ signal after the mentioned requirements is shown in figure 5.24 for the B^0_s and B^0 mass regions. The total efficiencies obtained for $B^0_s \to \mu^+\mu^-$, $B^0 \to \mu^+\mu^-$ and $B_s \to \mu^+\mu^-\gamma$ in both mass regions are shown in table 5.11, together with the corresponding ratios. The generator efficiencies are considered to be the same for all the channels. For $B^0 \to \mu^+\mu^-$, the efficiency has been considered the same as for $B^0_s \to \mu^+\mu^-$.

[8]Physics here referring to not taking into account trigger lines designed, among others, for luminosity or detector studies.

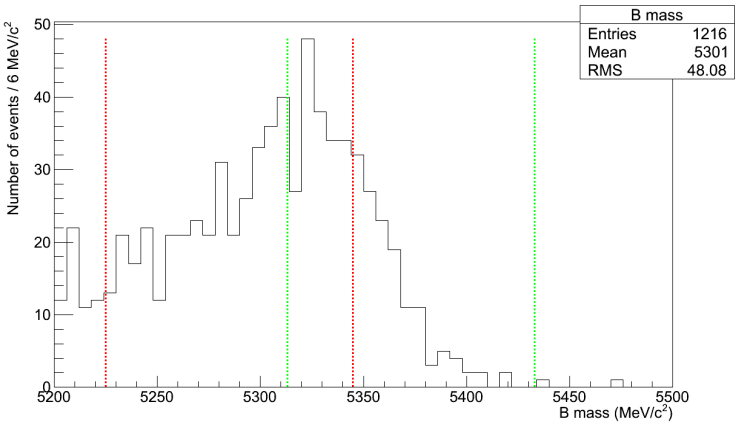

Figure 5.24: The invariant mass distribution of fake $B^0_{(s)} \to \mu^+\mu^-$ candidates from $B_s \to \mu^+\mu^-\gamma$ after selection cuts, trigger and BDT>0.8 (see text for details). Red vertical bars bound the B^0 mass region, while green vertical bars bound the B^0_s mass region.

	B^0_s	B^0
$\epsilon(B_s \to \mu^+\mu^-\gamma)$	$(0.0197 \pm 0.0011)\%$	$(0.0325 \pm 0.0014)\%$
$\epsilon(B_{d,s} \to \mu\mu)$	$(6.7475 \pm 0.0547)\%$	–
Efficiency ratio	0.0029 ± 0.0002	0.0048 ± 0.0002

Table 5.11: Total efficiencies of $B^0 \to \mu^+\mu^-$, $B^0_s \to \mu^+\mu^-$ and $B_s \to \mu^+\mu^-\gamma$ after selection, trigger and BDT>0.8 in B^0_s and B^0 mass regions (see text for details). The generator efficiency of the three channels is considered to be the same. For $B^0 \to \mu^+\mu^-$, the same efficiency than $B^0_s \to \mu^+\mu^-$ is assumed. The efficiency ratios are also shown.

These efficiency ratios can now be used to obtain the expected yields of $B_s \to \mu^+\mu^-\gamma$ in the B^0 and B^0_s mass regions after assuming a particular branching ratio for this decay. In this way, it must be said that the number of SM $B^0_s \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) candidates expected in the region of BDT>0.8 in the currently analysed data sample is ∼2.0 (∼0.2). These numbers come from the use of the normalisation factors that will be obtained in section 5.9. Using now the mentioned yields, the upper limit for the expected $B_s \to \mu^+\mu^-\gamma$ background in the B^0_s mass region is given by:

$$N(B_s \to \mu^+\mu^-\gamma) < \frac{\epsilon(B_s \to \mu^+\mu^-\gamma)}{\epsilon(B^0_s \to \mu^+\mu^-)} \times \frac{\mathcal{B}(B_s \to \mu^+\mu^-\gamma)}{\mathcal{B}(B^0_s \to \mu^+\mu^-)} \times N(B^0_s \to \mu^+\mu^-) \qquad (5.9)$$

with a very similar equation for the upper limits in the B^0 mass region. With all this, and assuming $\mathcal{B}(B_s \to \mu^+\mu^-\gamma) = 3.2 \times 10^{-8}$, the upper limits are of 0.058 (0.096) candidates in the B^0_s (B^0) mass region. These upper limits are smaller than the peaking background (as seen previously in this section) and almost negligible when compared to the combinatorial background extrapolated from the sidebands (as it will be seen in section 5.8). In fact, these limits are also small when compared to the errors in the combinatorial background expectations, so any potential contribution would be well covered by them.

In summary, the $B_s \to \mu^+\mu^-\gamma$ background has been checked to be under control for this level of statistics, even when its contribution may become more relevant in future searches, particularly for the $B^0 \to \mu^+\mu^-$ case.

5.7 The Boosted Decision Tree

The number of observed events after the selection is compared to the number of expected signal and background events in bins of two independent variables, the invariant mass and the output of a multivariate discriminant, the Boosted Decision Tree (BDT) constructed using the TMVA package [135]. Several other multivariate discriminants have been tested, the one chosen showing the better performance. The binning of the BDT and invariant mass distributions has also been optimised. This will be discussed in section 5.7.2.

5.7.1 BDT definition and performance

The BDT operator used in the $B^0_{(s)} \to \mu^+\mu^-$ analysis is built with the following variables: the B proper time (t), minimum impact parameter significance of the muons ($IPS(\mu)$), the impact parameter of the B ($IP(B)$), the distance of closest approach between the two muons (DOCA), the isolation of the two muons with respect to any other track in the event (μ isolation, $I(\mu)$), the transverse momentum of the B ($p_T(B)$), the cosine of the angle between the muon momentum in the B rest frame and the vector perpendicular to the B momentum and the beam axis ($\cos P$), the B isolation (CDF definition [145]) $I(B)$ and the minimum $p_{T,min}(\mu)$ of the two muons. The description of these variables is done in detail elsewhere [168]. Only two variables (DOCA and $IP(B)$) are in common with the BDTS discriminant (section 5.4.1).

The BDT training has been done using $B^0_s \to \mu^+\mu^-$ and $b\bar{b} \to \mu\mu X$ simulated events which have passed the cuts listed in table 5.3 including the BDTS cut but excluding the DLL cuts.

Moreover, the simulated events used for the training are required also to pass the trigger at each level[9]. This allows to train the BDT with events covering a similar phase space as real data. In this sense the trigger efficiency on $B^0_s \to \mu^+\mu^-$ selected events is ∼ 87% while on $b\bar{b} \to \mu\mu X$ selected events is ∼ 85%, so the trigger rejects about 15 % of both samples.

The first half of the MC samples is used to train the BDT and the second half is used to measure its performance. Then the two halves are swapped (the half previously used to train is used to test and vice-versa) and the process is repeated. The total performance is then obtained

[9]The condition *L0GlobalPhys & HLT1GlobalPhys & HLT2GlobalPhys* = True has to be satisfied.

by merging the two datasets. In this way, the full Monte Carlo statistics can be used to train the discriminant and measure the performance without risking to overtrain.

The background rejection versus signal efficiency curves were shown in figure 5.7 for different BDTS cuts applied at the selection level. As already discussed in section 5.4, the curve used in the current analysis is the one corresponding to BDTS>0.05.

Two other combinations of variables entering the BDT have been tested, one with 13 variables (BDT13) and the other with 19 variables (BDT19). The list of variables of these two new combinations (BDT13 and BDT19) is shown here below together with the one with 9 variables used as baseline. The variables in bold are in common with the BDTS.

- **BDT or BDT9 (baseline):** $IP(B)$, $t(B)$, **DOCA**, $I(B)$, $I(\mu_1)+I(\mu_2)$, $minIP\chi^2(\mu)$, $p_T(B)$, $\cos P$, $p_{T,min}(\mu)$;
- **BDT13:** $IP\chi^2(B)$, $t(B)$, **DOCA**, $I(B)$, $I(\mu_1)$, $I(\mu_2)$, $\chi^2/ndof$ of the secondary vertex ($\chi^2(B)/ndof$), cosine of the angle formed by the direction of the B momentum and the direction defined by the primary and secondary vertices ($DIRA$), B flight distance (BFD) and B flight distance significance ($BFDS$), $minIP(\mu)$, $\cos(P)$;
- **BDT19:** $IP\chi^2(B)$, $t(B)$, **DOCA**, $I(B)$, $I(\mu_1)$, $I(\mu_2)$, $\chi^2(B)/ndof$, **DIRA**, BFD, $BFDS$, $minIP(\mu)$, $\cos(P)$, azimuthal angle between the two muons (ϕ), the polar angle between the two muons (θ), the angle between the μ^+ direction in the rest frame of the B candidate and the direction of the *thrust momentum*[10] also boosted in the same frame ($otherBboost$), the angle between the B candidate momentum and the direction of the trust angle ($otherBangle$), $IP(B)$, $p_T(\mu)$.

The rejection versus efficiency curves for the three different combinations of variables entering in the BDT are shown in figure 5.25 without (left) and with (right) the BDTS cut BDTS>0.05. When the BDTS cut is not applied in the selection, BDT13 and BDT19 perform better than the baseline BDT. However, when the BDTS cut is applied, the three configurations are equivalent. This is due to the fact that most of the new variables introduced in the BDT13/BDT19 overlap with the variables entering the BDTS, so that most of the gain is already achieved at the previous stage. This is not true for the baseline BDT, which really profits from the discrimination introduced by the BDTS.

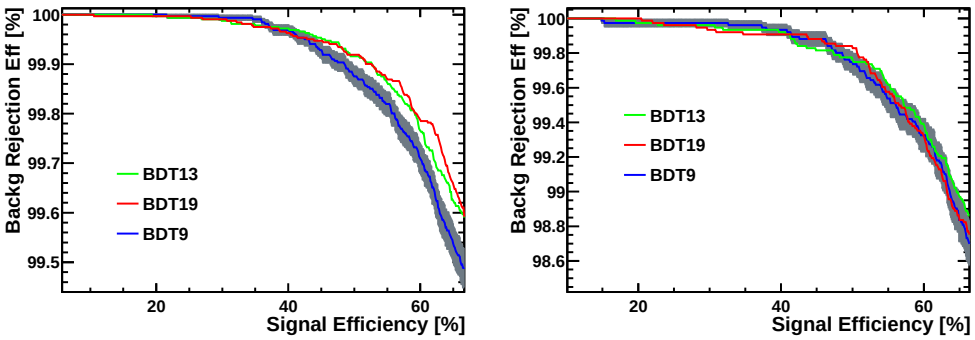

Figure 5.25: Rejection versus efficiency curves for the three different combinations of variables entering the BDT without (left) and with (right) the BDTS cut BDTS>0.05.

Finally, and to demonstrate the discrimination power of the variables entering in the baseline BDT, figure 5.26 shows the distributions of eight of these variables for signal $B_s^0 \to \mu^+\mu^-$ and $b\bar{b} \to \mu\mu X$ Monte Carlo events and for dimuon background events from data sidebands. The

[10]The *thrust* momentum should approximate the direction of the other B hadron for a signal event. It is defined as the sum of the momenta of all long tracks with IPS with respect to the PV associated to the B candidate between 4 and 40, $200\,\text{MeV}/c < p_T < 2000\,\text{MeV}/c$, $p < 30\,\text{GeV}/c$

last two distributions agree, again showing that the dominant background component in data is the $b\bar{b} \to \mu\mu X$ one.

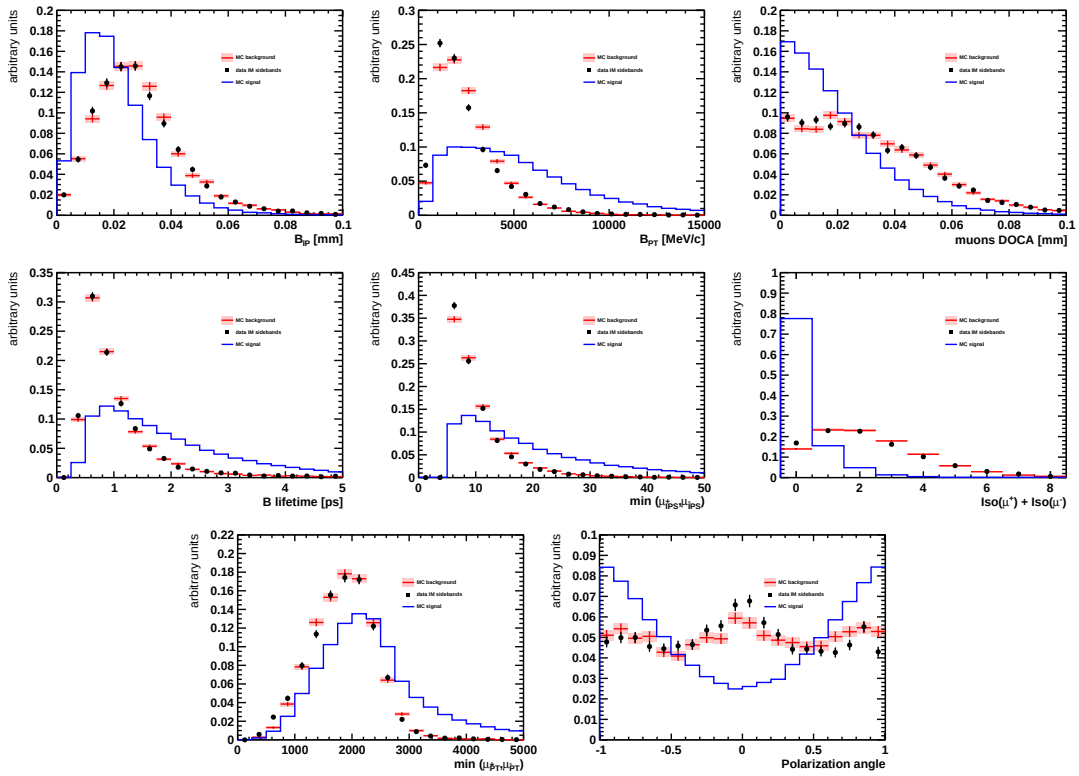

Figure 5.26: Distribution, for signal and background, of the variables used in the BDT definition. Blue: MC $B^0_{(s)} \to \mu^+\mu^-$ signal. Red: MC $b\bar{b} \to \mu\mu X$. Black, $B^0_{(s)} \to \mu^+\mu^-$ data sidebands.

5.7.2 Binning of the BDT and Invariant Mass distributions

The extraction of the limit is performed by doing a N-counting experiment, in which the information of the BDT and invariant mass distributions is combined using a modified frequentist approach or CL_s method. In this way, the 2-D space formed by these two variables is binned, which permits classifying the $B^0_{(s)} \to \mu^+\mu^-$ candidates surviving the selection.

The aim of this section is presenting the CL_s method and finding out which is the binning that maximises the sensitivity for the current analysis. This study is described in detail in [169].

5.7.2.1 The CL_s method

The compatibility of the distribution of events inside each bin of the search window in the invariant mass-BDT plane with a given branching fraction hypothesis is evaluated using the CL_s method. This method provides three estimators: CL_{s+b}, a measure of the compatibility of the observed distribution with the signal-plus-background hypothesis, CL_b, a measure of the compatibility with the background-only hypothesis and CL_s, a measure of the compatibility of the observed distribution with the signal-plus-background hypothesis normalised to the background-only hypothesis. These estimators are related to the the classifier used to do the test statistics Q defined as:

$$Q = \prod_{i=1}^{N} P(d_i, s_i(\mathcal{B}) + b_i)/P(d_i, b_i) \tag{5.10}$$

where the products runs over the number of bins, P is the Poisson distribution with expected values $s_i + b_i$ and b_i respectively, s_i and b_i denote the expected number of signal and background events in bin i and d_i is the number of observed events in bin i. The number of expected signal events depends on the assumed branching fraction and integrated luminosity, the number of expected background events and observed events depend on the integrated luminosity.

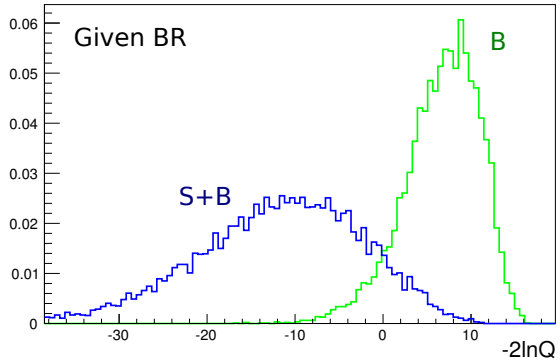

Figure 5.27: Distributions of -2ln Q for signal-plus-background hypothesis (Q_{s+b}) and background-only (Q_b) hypothesis.

The distributions of Q for the signal-plus-background hypothesis (Q_{s+b}) and background-only (Q_b) hypotheses are shown in figure 5.27. The quantity $\mathrm{CL_{s+b}}$ is the integral of the Q_{s+b} distribution from the observed value upwards, and hence represents the probability that another experiment gives a lower likelihood than the observed one under the hypothesis of signal-plus-background. $\mathrm{CL_{s+b}}$ is a measure of the incompatibility with the signal-plus-background hypothesis. The quantity $\mathrm{CL_b}$ is the integral of the Q_b distribution from the observed value upwards, hence the quantity 1-$\mathrm{CL_b}$ represents the probability that another experiment gives a lower likelihood than the observed one under the hypothesis of background-only. 1-$\mathrm{CL_b}$ is a measure of the incompatibility with the background hypothesis. $\mathrm{CL_s}$ is the ratio of confidence levels $\frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}}$, and is it typically the one used to set the exclusion (upper) limit on the branching fractions, whereas $1 - \mathrm{CL_b}$ is used as a $p - value$ to claim for an evidence or observation. The use of $\mathrm{CL_s}$ instead of $\mathrm{CL_{s+b}}$ intends to protect the result against negative statistical fluctuations of the background, which could potentially lead to the exclusion of the null hypothesis, even without any experimental sensitivity. In this way, a 95(90)% confidence level exclusion corresponds to $\mathrm{CL_s} = 0.05(0.1)$. A 3σ evidence corresponds to $1 - \mathrm{CL_b} = 2.7(1.35) \times 10^{-3}$ and a 5σ discovery to $1 - \mathrm{CL_b} = 5.73(2.87) \times 10^{-7}$ for one–sided (two—sided) definition of the significance.

The uncertainties on the input parameters are also taken into account by fluctuating the signal and background expectations according to the formula:

$$x_i' = x_i \times (1 + \frac{1}{2} r(s_+ - s_-) + \frac{1}{2} r^2 (s_+ + s_-)) \tag{5.11}$$

where x_i is the central value of the parameter, r is a random number generated from a Gaussian distribution of zero mean and unit width, and s_+ and s_- are the relative (signed) errors of x_i [170].

5.7.2.2 Optimal binning

The largest separation of the two distributions in figure 5.27 for a given signal yield (hence a given branching fraction hypothesis) can be used to produce a figure of merit to obtain the best sensitivity in a search. The separation is quantified by the variable:

$$\Delta LQ = 2 \ln Q_{s+b}^{\mathrm{med}} - 2 \ln Q_b^{\mathrm{med}} \tag{5.12}$$

where Q^{median}_{s+b} and Q^{median}_{b} are the medians of the Q distributions calculated in the signal plus background and background-only hypotheses. The separation between the two distributions as a function of the branching fraction hypothesis is shown graphically in figure 5.28: for a given branching fraction, the error around the median gives the 68% containement of possible results. Intuitively the more the two bands are separated the larger is the sensitivity.

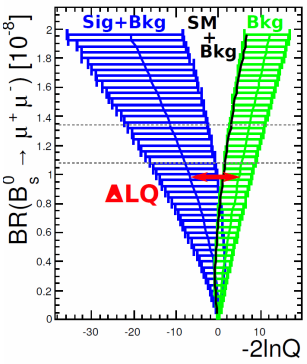

Figure 5.28: Separation of the medians of the Q_{s+b} and Q_b distributions as a function of the $B^0_s \to \mu^+\mu^-$ branching fraction. The errors around the medians give the 68% containement of possible results.

The choice of the binning is made out of a comparison of many set of bins on their ΔLQ. The binning is optimised for the $B^0_s \to \mu^+\mu^-$ decay, assuming an integrated luminosity of 1 fb^{-1} and the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ from Standard Model predictions.

The binning search is done in two steps: first the mass binning is optimised for events with BDT between 0.8 and 1.0. For a fixed number of bins, different configurations are tested and the pattern which gives the largest ΔLQ is recorded. In figure 5.29 top the best patterns corresponding to a given number of bins are shown together with the corresponding value of ΔLQ.

Nbin												ΔlnQ
6	5313	5331	5343	5355	5397	5415						6.478
6	5313	5331	5349	5391	5403	5415						6.478
7	5313	5331	5343	5355	5391	5403	5415					6.497
8	5313	5331	5343	5355	5391	5403	5409	5421				6.504
9	5313	5325	5337	5343	5355	5391	5403	5409	5421			6.510
10	5313	5325	5331	5337	5343	5355	5391	5403	5409	5421		6.514
10	5313	5325	5337	5343	5355	5391	5403	5409	5415	5421		6.514
11	5313	5325	5337	5343	5349	5361	5385	5397	5403	5409	5421	6.517

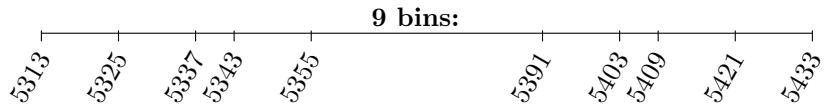

Figure 5.29: Optimal mass binning search. Top, best mass binnings with different number of bins. Bottom, best mass binning with 9 bins.

The ΔLQ value as a function of the number of mass bins is shown in figure 5.30 left. Given that after 9 bins the gain is very little, this is the option selected. The configuration with 9 bins that gives the larger ΔLQ is shown graphically in figure 5.29 bottom.

The optimal mass binning is then used to find the best binning configuration for the BDT. The ΔLQ as a function of the number of bins is shown in figure 5.30 right. The optimal 8 bins set reaches a ΔLQ value greater than 99% of the maximum possible value, so becomes the one selected.

As a summary, the best bin configuration is shown in table 5.12 both for mass and BDT.

The contribution of each bin to the analysis sensitivity can be evaluated using the additivity

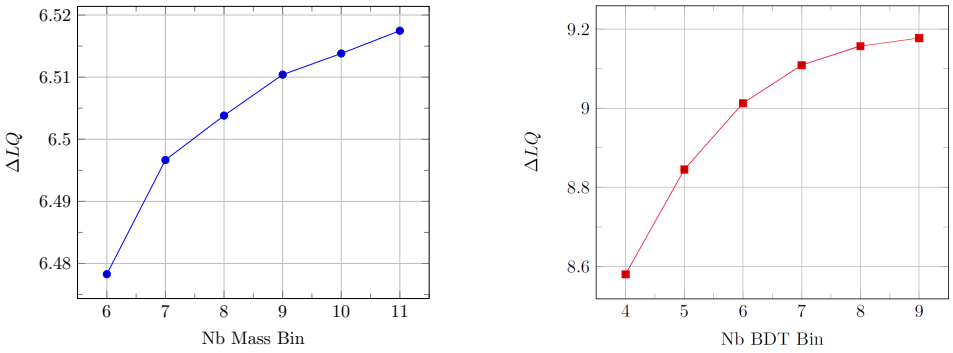

Figure 5.30: ΔLQ value as a function of the number of mass (left) and BDT (right) bins.

Bins	bin 1	bin 2	bin 3	bin 4	bin 5	bin 6	bin 7	bin 8	bin 9
Mass	5313–5325	5325–5337	5337–5343	5343–5355	5355–5391	5391–5403	5403–5409	5409–5421	5421–5433
BDT	0–0.25	0.25–0.4	0.4–0.5	0.5–0.6	0.6–0.7	0.7–0.8	0.8–0.9	0.9–1.0	

Table 5.12: Best bin configurations for mass and BDT.

of ΔLQ:

$$\Delta LQ = \sum_{bin} \Delta LQ_i$$

where ΔLQ_i is the ΔLQ in bin i. Figure 5.31 shows the contribution that each bin brings to the total ΔLQ and their ranking with respect to this criteria. Table 5.13 shows the contributions of each BDT bin if the mass bins are added up: more than 85 % of the total sensitivity is contained in the BDT bins above 0.7 while the first bin contributes less than 0.1%.

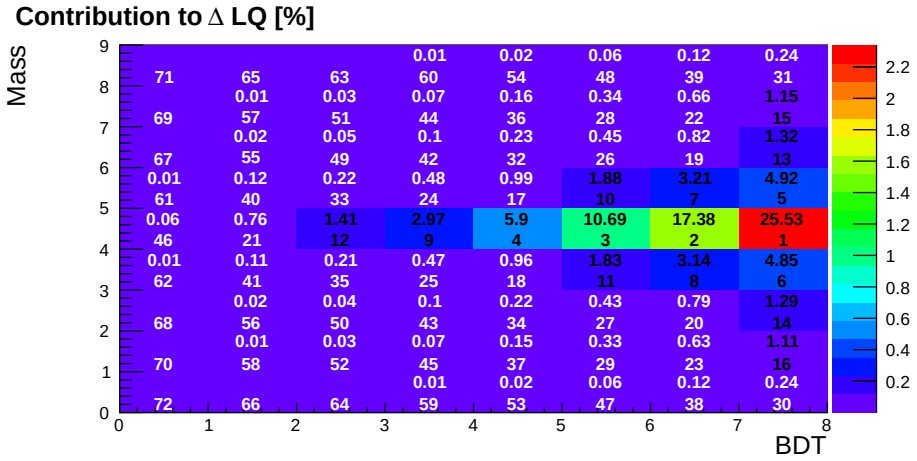

Figure 5.31: Contribution of each bin to the total ΔLQ and their ranking.

BDT bin	ΔLQ_i	$\Delta LQ_i/\Delta LQ$ (%)
0 – 0.25	0.00899	0.0981
0.25 - 0.4	0.1	1.09
0.4 – 0.5	0.186	2.03
0.5–0.6	0.396	4.33
0.6 – 0.7	0.797	8.70
0.7 – 0.8	1.48	16.1
0.8 – 0.9	2.47	26.9
0.9 – 1.0	3.73	40.7

Table 5.13: Absolute (second column) and relative (third column) contribution to the total ΔLQ for each BDT bin.

5.8 Calibration of the BDT and the invariant mass

The BDT discriminant is trained using Monte Carlo samples ($B^0_{(s)} \to \mu^+\mu^-$ for signals and $b\bar{b} \to \mu\mu X$ for background) and its PDF for signal and background events is extracted from data.

The BDT PDF for the signals is extracted using TIS events from the inclusive $B^0_{(s)} \to h^+h'^-$ sample (section 5.8.1), while the exclusive $B^0_{(s)} \to KK, K\pi, \pi\pi$ samples can be used to cross check the results.

The mass PDFs for the signals are parameterised as two Crystal Ball functions:

- the central values are extracted from the exclusive $B^0 \to K^+\pi^-$ and $B^0_s \to K^+K^-$ samples (section 5.8.2);
- the resolutions are measured by interpolating at the B^0 and B^0_s mass point the resolutions measured with the J/ψ, $\Psi(2S)$, $\Upsilon(1S, 2S, 3S)$ dimuon resonances (section 5.8.3);
- the turning point of the radiative tail is obtained by fitting the mass lineshape of $B^0_s \to \mu^+\mu^-$ Monte Carlo events smeared in such a way to recover the mass resolution measured in data.

The BDT and invariant mass PDFs for the combinatorial background are extracted from a fit of the dimuon mass sidebands in each BDT bin (section 5.8.4).

The $B^0_{(s)} \to h^+h'^-$ peaking background yield with double misidentified hadrons has been evaluated in section 5.6. Its mass lineshape is obtained from a Monte Carlo sample of double misidentified $B^0_{(s)} \to h^+h'^-$ events and normalised to the number of events expected in the two search windows. The BDT PDF of the double misidentified $B^0_{(s)} \to h^+h'^-$ is evaluated by convoluting the fake rate $\epsilon(K \to \mu)$ and $\epsilon(\pi \to \mu)$ probability extracted from the $D^0 \to K\pi$ decay to the (p,p_T) spectrum of selected $B^0_{(s)} \to h^+h'^-$ events in a given BDT bin. These two PDFs have been already discussed in section 5.6.3.

5.8.1 Extraction of the BDT for signal with the $B^0_{(s)} \to h^+h'^-$ inclusive sample

The BDT PDF for the $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ signals has been extracted from data using a sample of TIS[11] $B^0_{(s)} \to h^+h'^-$ inclusive events in different BDT bins and obtaining the number of signal candidates in each bin. The invariant mass line shape of $B^0_{(s)} \to h^+h'^-$ decays,

[11] in particular, TIS events at the L0 and HLT1 level have been considered, while the bias introduced by the HLT2 trigger level is computed from Monte Carlo and added as a correction to the BDT shape. This will be seen in section 5.9.

needed to extract yields, is presented in detail in [171]. In this case, three independent fits have been used, so that the maximum spread obtained between them is computed as a systematic error in the BDT PDF.

- The first is a 2D fit of the invariant mass lineshape (in $\mu\mu$ hypothesis) and of the momentum unbalance of the decay, $\beta = (p^+ - p^-)/(p^+ + p^-)$, where $p^\pm$ are the momenta of the tracks with positive/negative charge. The role of β in the fit is to disentangle among the different $B^0_{(s)} \to h^+h'^-$ decay modes.

- The second (1D, to invariant mass lineshape), uses a double Crystal Ball PDF for the signal and an exponential and Argus functions for the combinatorial and physical backgrounds.

- Finally, the third is also 1D, and uses a single Crystal Ball with radiative tail parameters fixed from simulation and resolution fixed from data for signal, while for the background the same parameterisations than in the second case are chosen.

The background is composed by a combinatorial component and a "physical background" due to the partial reconstruction of three-body B meson decays, like $B^0 \to \rho^\pm(\pi^0\pi^\pm)\pi^\mp$, in which a $\pi^+\pi^-$ pair mimics the $B^0_{(s)} \to h^+h'^-$ signal. Due to the missing π^0, the $\pi^+\pi^-$ invariant mass distribution is kinematically limited to the region below 5.14 GeV/c^2 (i.e. about $m_{B^0} - m_{\pi^0}$), apart from experimental resolution effects which might lead to higher values. For this reason, this background component mainly affects the left tail of the signal mass distribution. Other relevant decay modes belonging to this category are $B \to \rho K$ and $B \to K^*\pi$, as well as decays involving other intermediate resonances. Also, a small component due to $\Lambda_b^0 \to pK^-$ and $\Lambda_b^0 \to p\pi^-$ is present in the right sideband.

Further details on the fitting models, particularly on the first, can be found in [168, 172, 173].

Because of the distribution of the background BDT, to be seen in section 5.8.4, the first BDT bin is dominated by background. Because of this, the number of $B^0_{(s)} \to h^+h'^-$ events in this bin is evaluated by subtracting to the total yield of $B^0_{(s)} \to h^+h'^-$ TIS events the yields obtained in the seven higher BDT bins. As a cross–check, the yield of the first BDT bin is also independently fitted.

The fit projections for mass and β (using the first, 2D fit) for the total $B^0_{(s)} \to h^+h'^-$ sample and for the BDT bins above 0.25 are shown in figures 5.32 and 5.33.

The other two 1D fits give compatible results with the 2D one. In this way, figure 5.34 shows the invariant mass distributions for all BDT bins for the second model explained above, and table 5.14 compares the BDT PDFs obtained with the three models. The bin-by-bin maximum difference is included as a systematic error in the evaluation of signal BDT PDF.

BDT bin	$B^0_{(s)} \to h^+h'^-$ 2D-fit model 1 [%]	$B^0_{(s)} \to h^+h'^-$ 1D-fit model 2 [%]	$B^0_{(s)} \to h^+h'^-$ 1D-fit model 3 [%]
0 – 0.25	23.7 ± 3.6	24.4 ± 3.4	21.1 ± 3.7
0.25 – 0.4	11.2 ± 0.7	12.6 ± 0.7	12.5 ± 0.9
0.4 – 0.5	8.9 ± 0.5	9.8 ± 0.7	9.5 ± 0.7
0.5 – 0.6	8.6 ± 0.5	8.9 ± 0.5	9.1 ± 0.6
0.6 – 0.7	10.3 ± 0.6	10.0 ± 0.5	9.8 ± 0.6
0.7 – 0.8	11.0 ± 0.6	10.8 ± 0.6	11.1 ± 0.6
0.8 – 0.9	12.7 ± 0.7	12.2 ± 0.6	12.7 ± 0.7
0.9 – 1.0	13.4 ± 0.7	11.5 ± 0.7	14.2 ± 0.7

Table 5.14: Fractional yields for the eight BDT bins: 2D fit and 1D fit results, with statistical errors.

As for the mentioned problem in the first BDT bin, the comparison between both methods (total $B^0_{(s)} \to h^+h'^-$ yield minus yield in the BDT bins above 0.25 and direct fit) points to a

$\sim 4\%$ difference, which is added as an extra systematic error to this bin. As seen above, the contribution of the first bin to the limit is very small, so this uncertainty will not affect the final result.

The total yield of $B^0_{(s)} \to h^+h'^-$ events obtained is $N_{B^0_{(s)}\to h^+h'^-} = (16734 \pm 1312)$. This result includes a systematic error accounting for some effects in the the parameterisation of the β PDF of the 2D fitting model, for the change of the background shape depending on the BDT and BDTS cuts applied and also for the 5% maximum spread observed between the three fitting models. The $B^0_{(s)} \to h^+h'^-$ total yield is used not only in the BDT calibration but also in the computation of the normalisation factor for the $B^0 \to K^+\pi^-$ channel (see section 5.9) and in the evaluation of the total yield of the $B^0_{(s)} \to h^+h'^-$ events with double misID (as seen in section 5.6).

As explained before, and in order to keep a reasonable dataset for the calibration, the $B^0_{(s)} \to h^+h'^-$ events are required to be TIS events only at L0 and HLT1 stages and then corrected for possible HLT2 trigger biases. Therefore the BDT response calibrated with $B^0_{(s)} \to h^+h'^-$ TIS events has to be divided by the following correction factor:

$$\frac{\epsilon_{\rm TIS}(L0 \times HLT1)_{hh} \times \epsilon_{\rm trigger}(HLT2)}{\epsilon_{\rm trigger}(L0 \times HLT1 \times HLT2)_{\mu\mu}} \tag{5.13}$$

where the numerator is the efficiency correction to be applied to $B^0_{(s)} \to h^+h'^-$ TIS events to get a trigger unbiased sample, while the denominator is the efficiency correction to be applied to the unbiased sample to emulate the muon trigger bias on $B^0_{(s)} \to \mu^+\mu^-$ signal events.

The evaluation of the above factors has been performed using $B^0 \to K^+\pi^-$ and $B^0_{(s)} \to \mu^+\mu^-$ Monte Carlo events where the 2011 trigger has been emulated. The trigger bias as a function of the BDT output is shown in figure 5.35, on left plot for the hadronic triggers, *i.e.* the numerator of equation 5.13, and on right plot for muon triggers, *i.e.* the denominator of equation 5.13. Both the hadronic and muon trigger biases have been evaluated on offline selected events. The results are shown in table 5.15 as a function of the BDT bins.

BDT bin	Trigger correction [%]
0.0-0.25	5.4±0.1
0.25-0.40	5.2 ±0.1
0.40-0.50	5.2±0.1
0.5-0.6	5.7±0.2
0.6-0.7	5.9±0.2
0.7-0.8	6.1±0.2
0.8-0.9	6.2±0.2
0.9-1.0	6.3±0.2

Table 5.15: Trigger corrections as a function of the BDT bins.

As shown in sections 5.5 and 5.6, the double misID background rejection greatly benefits from the introduction of a loose DLL selection cut, at a cost of $\sim 3\%$ loss of signal events. Even though this effect is small compared to other effects previously discussed, the behaviour of the DLL cut efficiency on signal, ϵ_{DLL}, as a function of BDT has been studied. The efficiency has been evaluated by convoluting the muon DLL efficiencies from data control samples with the MC signal events (section 5.5). The results are listed in table 5.16. The observed bias (*i.e.* the efficiency per bin normalised to its average value) has been introduced in the definition of the signal BDT PDF.

Finally, in figure 5.36, the calibrated BDT shape is shown before (left) and after (right) the corrections are applied. In order to properly compare the probability per bin, in the figure after the corrections (right hand side) a normalisation to the bin size is also included.

BDT bin	ϵ_{DLL} [%]
0.0-0.25	96.86 ± 0.01
0.25-0.4	96.86 ± 0.01
0.4-0.5	96.93 ± 0.01
0.5-0.6	97.10 ± 0.01
0.6-0.7	97.40 ± 0.01
0.7-0.8	97.76 ± 0.01
0.8-0.9	98.12 ± 0.01
0.9-1.0	98.54 ± 0.01

Table 5.16: DLL cut efficiency as a function of the BDT bins.

5.8.2 Invariant mass central values for signal with the $B^0_{(s)} \to h^+h'^-$ exclusive samples

The exclusive $B^0_{(s)} \to \pi K, \pi\pi, KK, K\pi$ channels can be used in data to estimate the BDT PDF and the B^0_s and B^0 masses. However, due to a still not yet fully understood control of the RICH calibration, the exclusive channels have been used to estimate the central mass values and only to cross-check the mass resolution and the BDT calibration. Because of this, only the first estimation is explained here, while the other two checks can be seen in [173].

As explained in section 3.2.3.3.1, the information of the Cherenkov angle and the number of photo-electrons produced by a given particle allows to perform a PID hypothesis test, via a delta-log-likelihood (DLL). In the kaon and pion hypothesis, a $DLL_{K-\pi}$ is built and a particle is identified with a certain probability $P(k)$ as a kaon if $DLL_{K-\pi} > k$ or as a pion if $DLL_{K-\pi} < -k$.

Using a $D^* \to D^0(K\pi)\pi$ sample (as in the case of the muon misID), a RICH efficiency map has been evaluated for the 2011 data. For each $DLL_{K-\pi}$ cut, an efficiency value is provided for each bin of the 3-dimensional phase space defined by the three axes ($p, \eta, N_{\text{tracks}}$[12]).

Exclusive $B^0_{(s)} \to h^+h'^-$ decays are extracted from $B^0_{(s)} \to h^+h'^-$ inclusive TIS events, where a loose cut in the BDT (BDT$>$ 0.2) is applied to cleanup part of combinatorial background[13]. A particle is consider a kaon if $DLL_{K-\pi} > 10$ (and $DLL_{K-p} > 0$) and a pion if $DLL_{K-\pi} < -10$ (and $DLL_{\pi-p} > 0$). From the original $B^0_{(s)} \to h^+h'^-$ sample, four disjunctive sets of events for different mass hypotheses $\pi^+\pi^-$, $K^+\pi^-$, π^+K^- and K^+K^- are built-up. The distribution of the $B^0_{(s)} \to h^+h'^-$ exclusive decays mass is obtained weighting each event by the DLL-cut efficiency. For a given $DLL_{K-\pi}$ cut k, the weight used is given by:

$$w(k) = \frac{1}{\epsilon_h(p(h^+)), \eta(h^+), N_{\text{tracks}} || DLL_{K-\pi}| > k} \cdot \frac{1}{\epsilon_h(p(h^-)), \eta(h^-), N_{\text{tracks}} || DLL_{K-\pi}| > k} \tag{5.14}$$

The mass distributions are shown in figure 5.37 for the K, π hypothesis (left: π^+K^-, right: $K^+\pi^-$) , and in figure 5.38 left for the K^+K^- hypothesis and figure 5.38 right for the $\pi^+\pi^-$ one. The mass distribution is fitted with the second model (1D) seen in section 5.8.1: a double Crystal Ball function for the signal, an Argus function for the physical background and an exponential function for the combinatorial background. The values of the mass average for the four combinations are listed in table 5.17. Notice that there is a bias of 0.1 % above the values in [5]. These values include a systematic error, estimated varying the PID and BDT cuts and performing the fit fixing or not the tail parameter to the MC value.

[12] Total number of tracks in the event, to take care of the dependence of the PID on the event multiplicity.

[13] A BDT cut does not affect the mass resolution as the BDT output and mass resolution are fully uncorrelated, see [168].

$B^0 \to \pi^+\pi^-$	$5283.98 \pm 0.63_{(\text{stat})} \pm 0.21_{(\text{syst})}$
$B^0 \to K^+\pi^-$	$5285.27 \pm 0.29_{(\text{stat})} \pm 0.32_{(\text{syst})}$
$B^0 \to \pi^+K^-$	$5284.31 \pm 0.32_{(\text{stat})} \pm 0.25_{(\text{syst})}$
B^0 combined	$5284.63 \pm 0.20_{(\text{stat})} \pm 0.27_{(\text{syst})}$
$B^0_s \to K^+K^-$	$5372.96 \pm 0.32_{(\text{stat})} \pm 0.28_{(\text{syst})}$

Table 5.17: Mass of the exclusive $B^0_{(s)} \to h^+h'^-$ samples (in units of MeV/c^2).

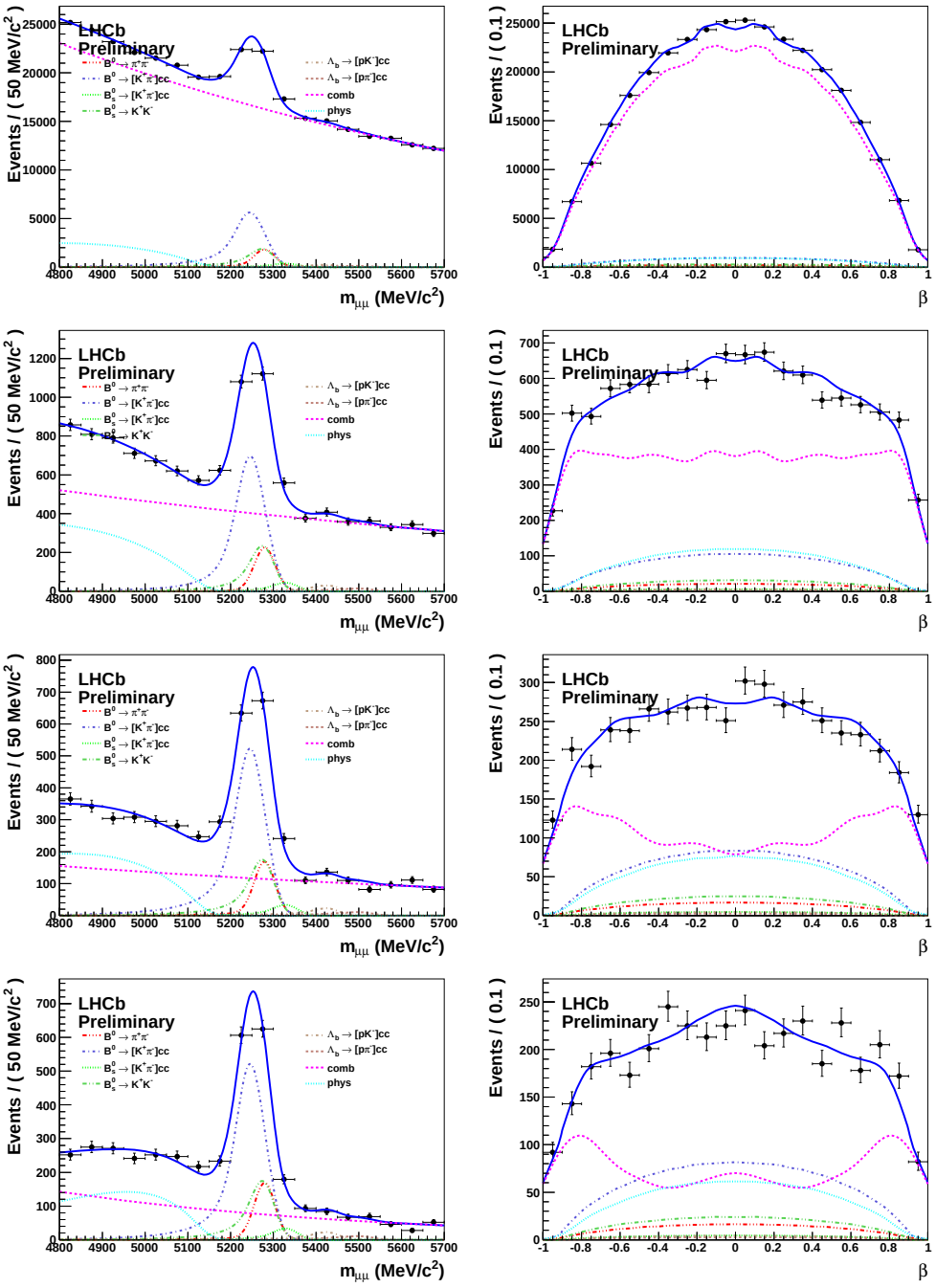

Figure 5.32: Fit projections on mass (left) and β (right): from top to bottom, whole BDT range, and BDT bins 2-3-4.

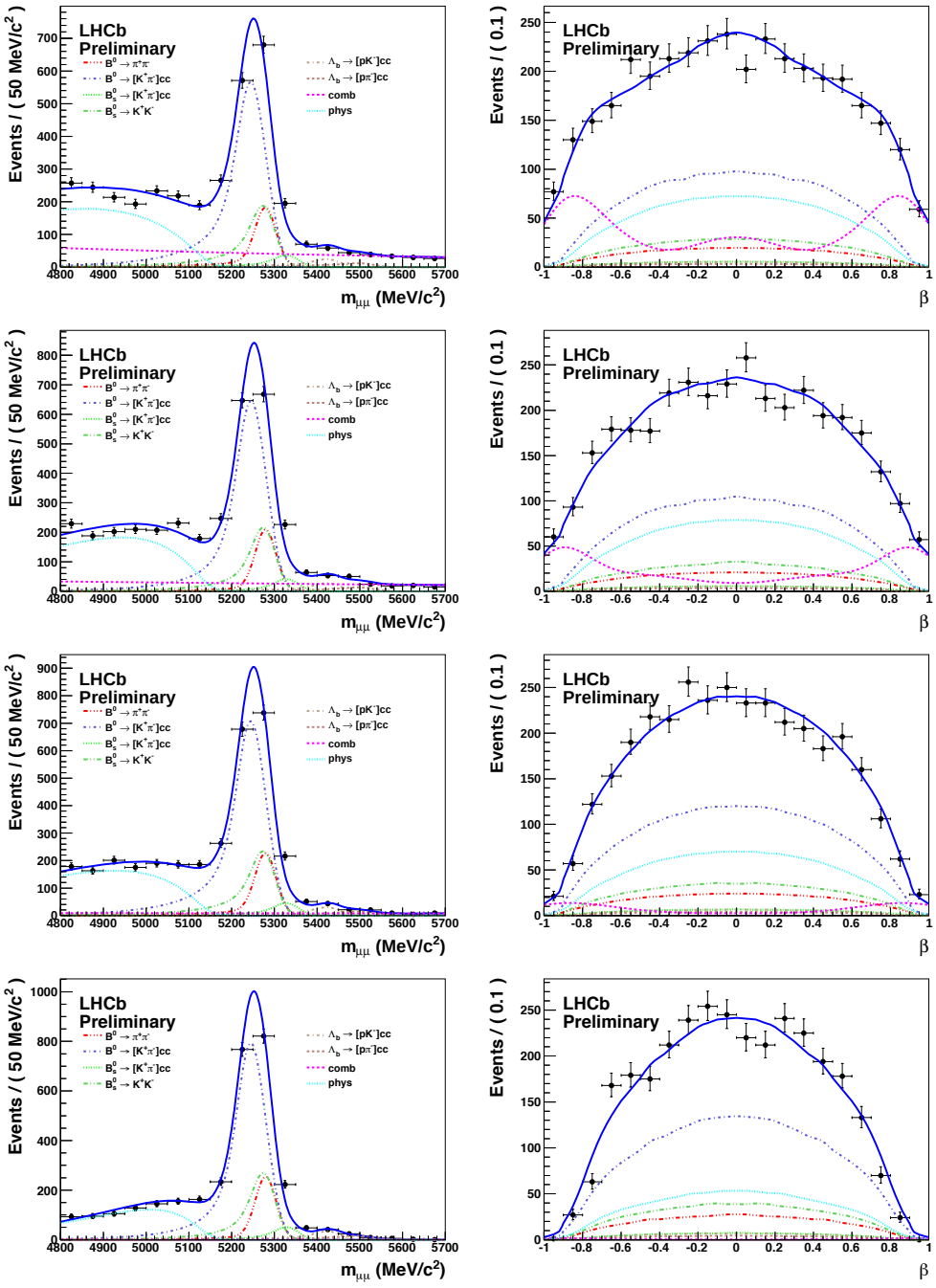

Figure 5.33: Fit projections on mass (left) and β (right): from top to bottom, BDT bins 5-6-7-8.

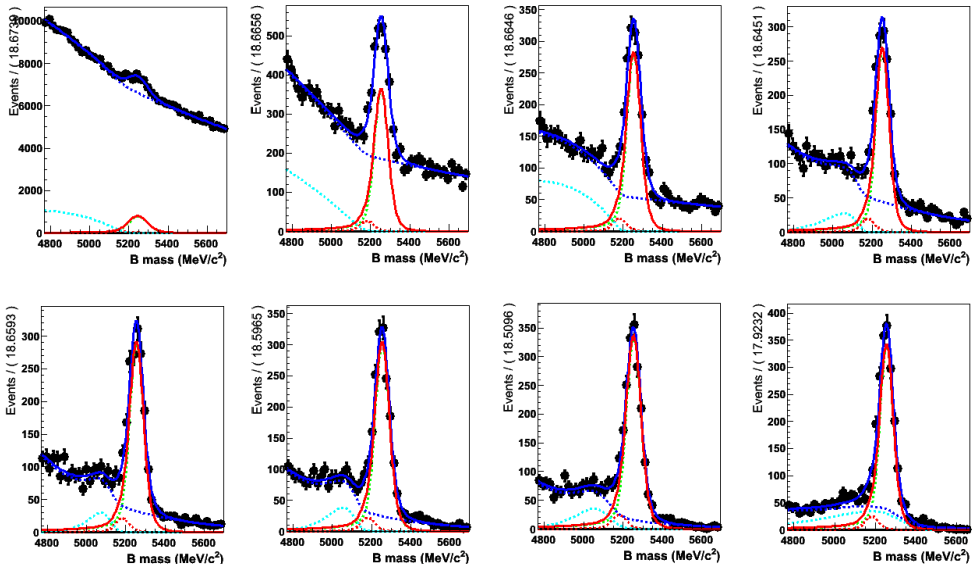

Figure 5.34: Mass fit results for all of the BDT bins with first 1D fit. The 10 bins are sorted from left to right, from top to bottom.

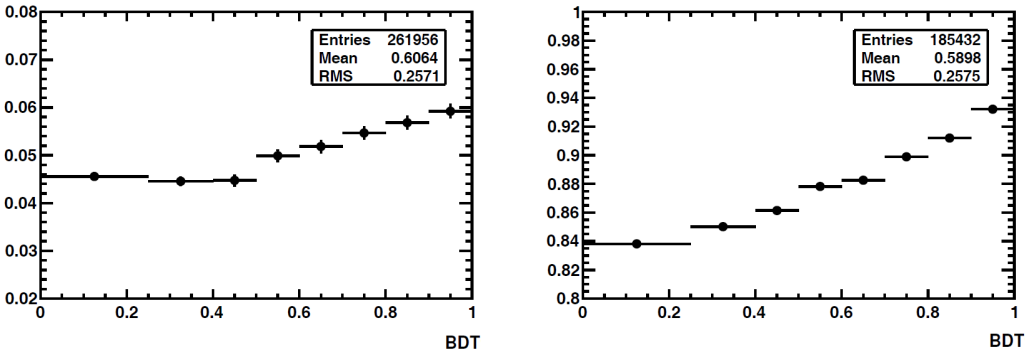

Figure 5.35: Trigger bias as a function of the BDT output. Left: bias due to the hadronic trigger. Right: bias due to the muon trigger. The y axis is in arbitrary units.

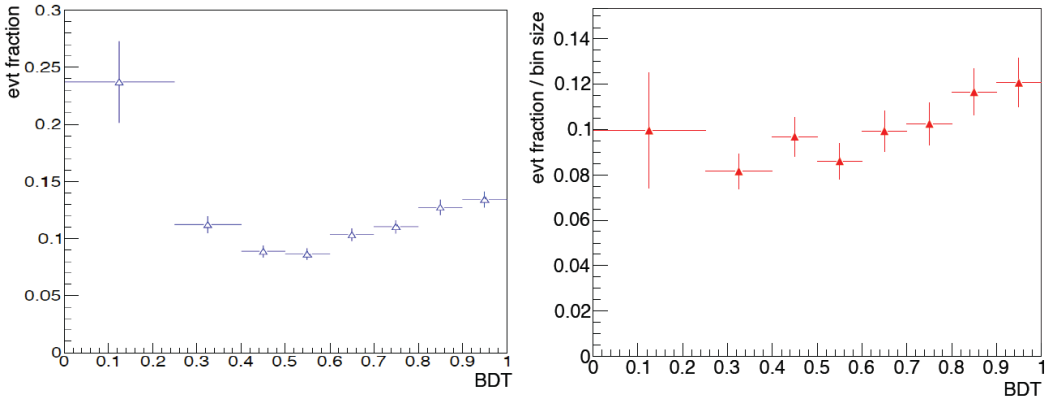

Figure 5.36: Calibrated BDT shape before (left) and after (right) the corrections and the normalisation to the bin size.

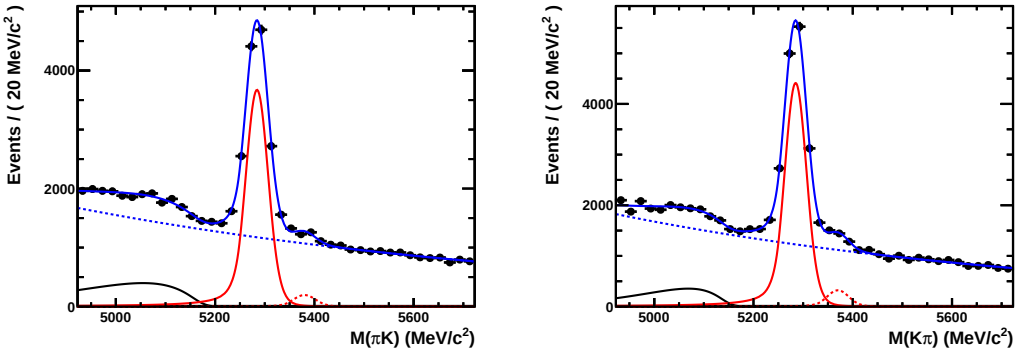

Figure 5.37: Mass distributions for $K\pi$ combinations (left: π^+K^-, right: $K^+\pi^-$).

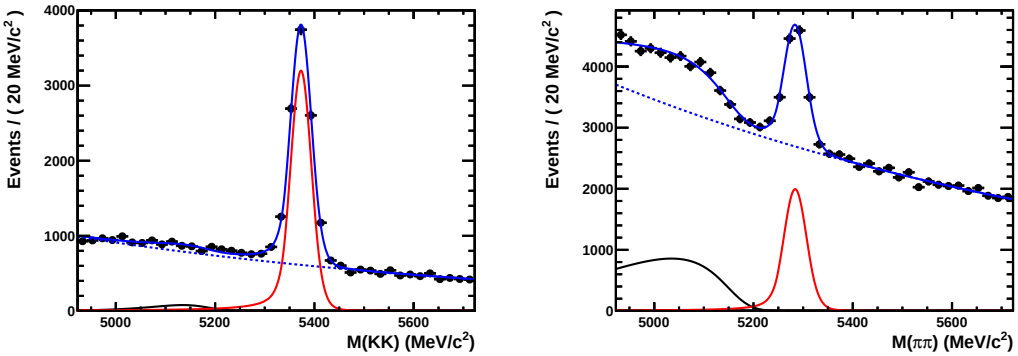

Figure 5.38: Mass distributions for K^+K^- (left) and $\pi^+\pi^-$ (right) combinations.

5.8.3 Calibration of the invariant mass resolution for signal using the interpolation method

The $m_{B_s^0}$ and m_{B^0} mass resolutions are extracted from data by performing an interpolation from the measured resolutions of Charmonium and Bottomonium resonances decaying significantly into two muons (i.e. J/ψ, $\psi(2S)$ and $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$).

The Υ resonances have been extracted with the *Dimuon High Mass* stripping line, the J/ψ from the *Jpsi2MuMuLine* (that selects the mass region $|m_{\mu\mu} - m_{J/\psi}| < 80$ MeV/c^2) and the $\Psi(2S)$ from the *Psi2MuMuLine* (that selects the mass region $|m_{\mu\mu} - m_{\Psi(2S)}| < 100\,\text{MeV}/c^2$). Some additional cuts, $p_T(\mu) > 1$ GeV, $\chi^2_{\text{track}}/ndof < 5$ and $\chi^2_{\text{PV}}/ndof < 10$ have been added to the stripping selections.

The events in the data samples are weighted such that the momentum spectra of the Charmonium and Bottomonium resonances are adjusted to the momentum spectrum of MC selected $B_s^0 \to \mu^+\mu^-$ events.

The mass lineshapes for charmonium and bottomonium resonances are shown in figure 5.39. The resonances are fitted with a double-sided Crystal Ball PDF (being the tail parameters fixed by Monte Carlo) with common resolution. The background is fitted with an exponential function.

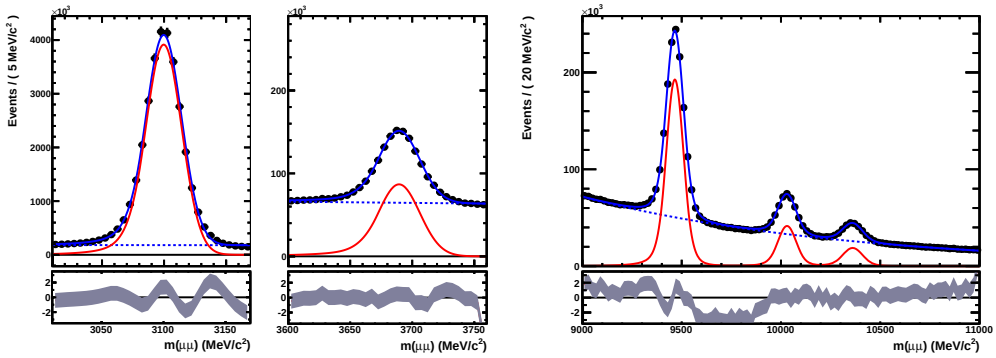

Figure 5.39: Dimuon invariant mass spectrum between (2.9 - 3.9) GeV/c^2 (left) and (9-11) GeV/c^2 (right) range. Bottom plots show the difference between the fitted PDF and the observation.

An interpolation between all these resonances to the B^0 and B_s^0 mass is then done. This has been performed using a power–law function:

$$\sigma_{\mu\mu}(m_{\mu\mu}) = a_0 + a_1 \cdot m_{\mu\mu}^{\gamma} \quad \gamma = 1.37 \pm 0.08 \tag{5.15}$$

The behaviour of the mass resolution in the extended mass range from the J/ψ to the Z^0 is shown in figure 5.40 (left): the power-law function (blue curve) nicely predicts the measured Z^0 mass resolution within 0.75 σ, giving a better result than a simple linear function, which is also shown in the figure (black curve).

The zoomed range between J/ψ and $\Upsilon(3S)$ is shown in figure 5.40 (right) where both the power-law and the linear fits are performed: the fit with a power-law function gives a $\chi^2/ndof = 3.7/2$ while the fit with a linear function $\chi^2/ndof = 16.3/3$. The power-law function clearly reproduces better the data also in the restricted range between J/ψ and $\Upsilon(3S)$. In table 5.18 the breakdown of the different components of the systematic errors is detailed for the two fits.

Finally the $m_{B_s^0}$ and m_{B^0} mass resolutions are as follows:

$$\begin{aligned} \sigma(B_s^0) &= (24.8 \pm 0.3_{\text{stat}} \pm 0.7_{\text{syst}})\,\text{MeV}/c^2 \\ \sigma(B^0) &= (24.3 \pm 0.3_{\text{stat}} \pm 0.6_{\text{syst}})\,\text{MeV}/c^2 \end{aligned} \tag{5.16}$$

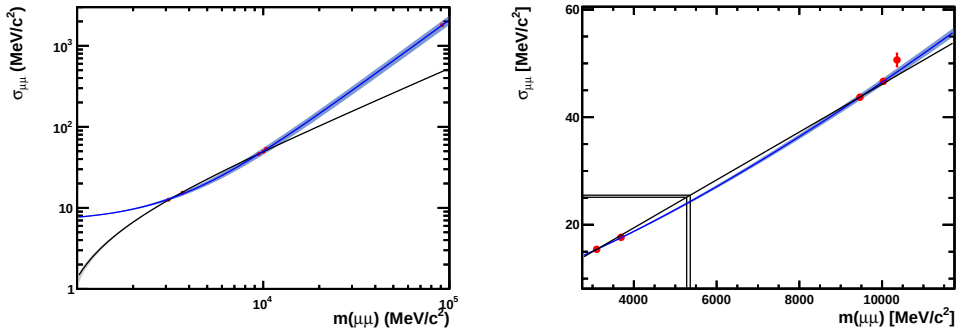

Figure 5.40: Left: behaviour of mass resolution in the mass range from J/ψ to the Z^0 with power-law (blue) and linear function (black) superimposed; right: zoom between J/ψ and $\Upsilon(3S)$. In the right plot, the position of the B^0 and B^0_s masses (and corresponding resolutions) is also shown.

Table 5.18: Mass resolution: breakdown of various contributions to the systematic error.

Systematics	B^0	B^0_s
Selection cuts	± 0.22 MeV/c^2	$\pm$ 0.18 MeV/c^2
Momentum weighting	$\pm$ 0.25 MeV/c^2	$\pm$ 0.25 MeV/c^2
Error asymmetry	$\pm$ 0.21 MeV/c^2	$\pm$ 0.26 MeV/c^2
Mass window	$\pm$ 0.32 MeV/c^2	$\pm$ 0.32 MeV/c^2
Fit function of invariant mass	$\pm$ 0.39 MeV/c^2	$\pm$ 0.46 MeV/c^2
Total systematic error	± 0.64 MeV/c^2	$\pm$ 0.69 MeV/c^2

where the systematic uncertainties have been added in quadrature. These results are in agreement with the method that uses the exclusive $B^0_{(s)} \to h^+h'^-$ samples, as it can be seen in [173].

5.8.4 BDT and invariant mass PDFs for combinatorial background

The BDT and invariant mass PDFs for the combinatorial background inside the signal regions are extracted from data by interpolating the number of expected events using the invariant mass sidebands for each BDT bin. In figure 5.41 the invariant mass distribution for events in each BDT bin is shown and the single exponential function used to interpolate the expected number of background events in the signal regions is superimposed.

The sidebands considered for the fit are defined by $m_{\mu\mu} = [4900\text{–}5000]\,\text{MeV}/c^2$ and $m_{\mu\mu} = [5433\text{–}6000]\,\text{MeV}/c^2$. The boundaries in the lower sidebands are chosen to exclude mass regions polluted by background sources other than combinatorial: cascading $b \to c\mu \to \mu\mu X$ decays below $4900\,\text{MeV}/c^2$ and $B^0_{(s)} \to h^+h'^-$ events with both hadrons misidentified as muons above $5000\,\text{MeV}/c^2$. The former contribution does not reach the signal region and consequently the fit remains unaffected by it. The latter contribution is modeled independently and should not affect the fit of the combinatorial background either. The lower boundary of the upper sideband is defined by the measured B^0_s mass plus $60\,\text{MeV}/c^2$.

The expected number of combinatorial background events in the B^0 and B^0_s mass windows as well as in the full blinded region (more precisely the region defined as the union of the B^0_s and the B^0 signal region) and the indices of the exponential functions are shown in table 5.19, column labeled 'blind'. Note that due to the insufficient number of events in the sidebands, the fit of the last bin of BDT would be meaningless. Thus the fit is performed to the events with

$0.8 < \mathrm{BDT} < 1.$ and the index of the resulting exponential is then used as constant input to fit the two last bins independently.

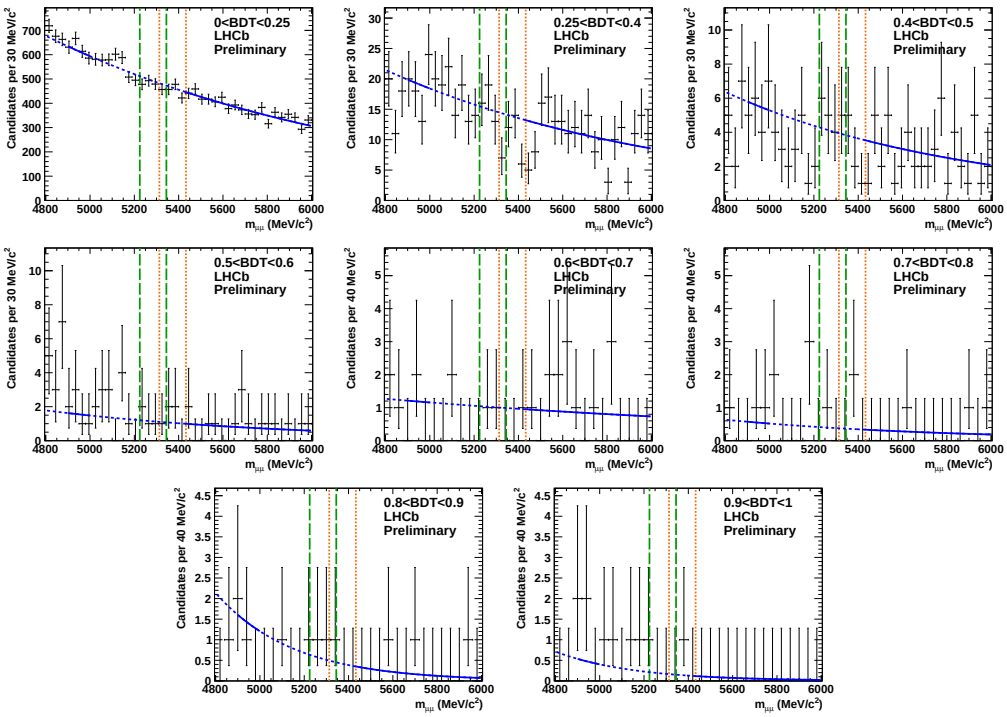

Figure 5.41: Distribution of the $\mu^+\mu^-$ invariant mass for events in each BDT output bin. The curve shows the model used to fit the sidebands and extract the expected number of combinatorial background events in the B_s^0 and B^0 signal regions, delimited by the vertical dotted orange and dashed green lines respectively. Only events in the region in which the line is solid have been considered in the fit.

In table 5.19 the results obtained with a double exponential function fitted to the extended invariant mass range [4173–5000] MeV/c^2 and [5433–6573] MeV/c^2 and with a linear fit in the nominal mass sidebands are also shown. A systematic uncertainty is introduced if the yields in the blinded regions differ by more than 1σ between the fit models. This is the case for the three first bins and the additional relative systematic are 1.8 %, 17.6 % and 21.8 %, respectively.

The final shapes for signal and background BDT PDFs are shown in figure 5.42.

Bin	exponential in [4900,5000] and [5433,6000]GeV k $[10^{-4}]$	B_s	B_0	Union of search windows	double exponential in [4173,5000] and [5433,6573] GeV Union of search windows
1	$-6.62^{+0.30}_{-0.30}$	1877^{+22}_{-21}	1990^{+21}_{-21}	3356^{+35}_{-35}	3295^{+23}_{-23}
2	$-7.7^{+1.8}_{-1.7}$	$55.9^{+3.8}_{-3.6}$	$59.9^{+3.7}_{-3.6}$	$100.5^{+6.3}_{-6.0}$	$82.9^{+3.8}_{-3.7}$
3	$-9.3^{+3.4}_{-3.3}$	$15.1^{+2.0}_{-1.8}$	$16.4^{+2.0}_{-1.9}$	$27.3^{+3.4}_{-3.1}$	$21.5^{+2.0}_{-1.9}$
4	$-9.2^{+6.5}_{-6.2}$	$4.25^{+1.08}_{-0.93}$	$4.61^{+1.11}_{-0.95}$	$7.7^{+1.8}_{-1.6}$	$7.7^{+1.2}_{-1.1}$
5	$-4.6^{+8.0}_{-7.4}$	$2.96^{+0.94}_{-0.80}$	$3.08^{+0.87}_{-0.73}$	$5.3^{+1.5}_{-1.2}$	$4.50^{+0.82}_{-0.72}$
6	-10^{+13}_{-12}	$1.07^{+0.58}_{-0.43}$	$1.17^{+0.61}_{-0.45}$	$1.95^{+1.01}_{-0.75}$	$1.37^{+0.39}_{-0.27}$
7	$-28.6^{+9.6}_{-10.4}$	$1.26^{+0.59}_{-0.45}$	$1.63^{+0.76}_{-0.58}$	$2.51^{+1.17}_{-0.89}$	$1.52^{+0.54}_{-0.40}$
8	$-28.6^{+9.6}_{-10.4}$	$0.42^{+0.37}_{-0.23}$	$0.54^{+0.48}_{-0.30}$	$0.84^{+0.74}_{-0.46}$	$0.45^{+0.25}_{-0.13}$

Table 5.19: Expected yields in the B^0_s and B^0 mass windows as well as in the union of those for different fit models and per bin of BDT. The indices of the exponential function are shown in the case of the single exponential model.

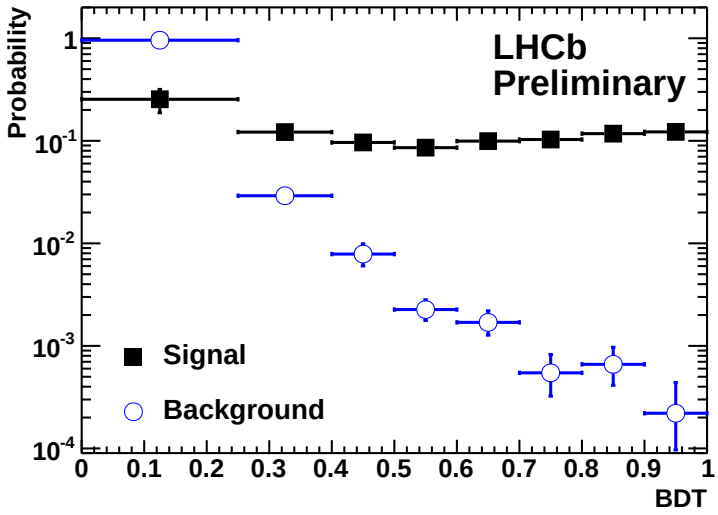

Figure 5.42: BDT PDFs for signal and background.

5.9 Normalisation

In order to estimate the signal branching ratio (as explained in section 4) the number of observed signal events has to be normalised to the number of events of a calibration channel with a well known branching ratio. In the case of $B^0_{(s)} \to \mu^+\mu^-$, three complementary normalisation channels have been used: $B^+ \to J/\psi(\mu^+\mu^-)K^+$, $B^0_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ and $B^0 \to K^+\pi^-$. The first two decays have similar trigger and muon identification efficiencies as the signal but different number of particles in the final state, while the third channel has very similar topology but very different trigger.

The selection of the normalisation channels, described in section 5.4, has been designed to be very similar to the selection of the signal events, such that the systematic uncertainties cancel in the ratio of efficiencies and the knowledge of the absolute integrated luminosity and the $b\bar{b}$ cross-section are not needed.

To translate the number of observed events into a branching ratio, the following equation is

used:

$$\mathrm{BR} = \mathrm{BR}_{\mathrm{cal}} \times \frac{\epsilon^{\mathrm{REC}}_{\mathrm{cal}} \epsilon^{\mathrm{SEL|REC}}_{\mathrm{cal}} \epsilon^{\mathrm{TRIG|SEL}}_{\mathrm{cal}}}{\epsilon^{\mathrm{REC}}_{\mathrm{sig}} \epsilon^{\mathrm{SEL|REC}}_{\mathrm{sig}} \epsilon^{\mathrm{TRIG|SEL}}_{\mathrm{sig}}} \times \frac{f_{\mathrm{cal}}}{f_{B^0_q}} \times \frac{N_{B^0_{(s)} \to \mu^+\mu^-}}{N_{\mathrm{cal}}} = \alpha_{\mathrm{cal}} \times N_{B^0_{(s)} \to \mu^+\mu^-} \,, \quad (5.17)$$

where $f_{B^0_q}$ and f_{cal} are the probabilities that a b-quark fragments into a B^0_q and into the b-hadron relevant for the chosen calibration mode. $\mathrm{BR}_{\mathrm{cal}}$ is the branching ratio and N_{cal} is the number of selected events of the calibration channel. The efficiency is separated in three factors: ϵ^{REC} is the efficiency to reconstruct all the tracks of the decay including the geometrical acceptance of the detector; $\epsilon^{\mathrm{SEL|REC}}$ is the efficiency to select the events which have been reconstructed; $\epsilon^{\mathrm{TRIG|SEL}}$ is the efficiency of the trigger on reconstructed and selected events. The sub-indexes indicate if the efficiency refers to the signal (sig) or the calibration channel (cal). Finally, α_{cal} is the normalisation factor (or single event sensitivity) and $N_{B^0_{(s)} \to \mu^+\mu^-}$ the number of observed signal events.

This section is devoted to the estimation of the factors that enter in the computation of α_{cal}.

5.9.1 Ratio of production fractions

The ratio $f_{B^0}/f_{B^0_s} = f_{B^+}/f_{B^0_s}$ has been recently determined by LHCb in two different ways: using the relative abundance of $B^0_s \to D^-_s\pi^+$, $B^0 \to D^-K^+$ and $B^0 \to D^-\pi^+$ [174] and using semileptonic $B \to DX$ decays [175]. The average of the two LHCb results is $f_s/f_d = 0.267^{+0.021}_{-0.020}$ [176]. This number will be used in the limit computation.

5.9.2 Branching ratios of the control channels

Table 5.20 shows the values of the $\mathcal{B}$ of the three control channels. The error on the $\mathcal{B}$ of $B^0_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ is 26% and is the dominant uncertainty when this channel is used for the normalisation.

$B^+ \to J/\psi K^+$	$B^0_s \to J/\psi\,\phi$	$B^0 \to K^+\pi^-$
$(6.01 \pm 0.21) \times 10^{-5}$	$(3.4 \pm 0.9) \times 10^{-5}$	$(1.94 \pm 0.06) \times 10^{-5}$

Table 5.20: Branching ratios of channels used for the normalisation

5.9.3 The ratio of reconstruction efficiencies

The acceptance and reconstruction efficiencies depends on the phase space of the final decay products, the acceptance of the detector and the efficiency of track finding algorithms. The ratio of acceptance and reconstruction efficiencies is computed using the MC simulation. The differences between the simulation and data are treated as systematic uncertainties.

The factors entering in the estimation of ϵ^{REC} are described next.

5.9.3.1 Reconstruction efficiencies from MC simulation

The reconstruction efficiency factorises in two parts: the detector acceptance (or generation efficiency) ϵ_{gen} and the reconstruction efficiency $\epsilon_{\mathrm{reco/gen}}$. The detector acceptance is defined as the fraction of the tracks in the final state that are inside the LHCb acceptance[14]. The reconstruction efficiency is defined as the efficiency to reconstruct all the tracks in the final state.

[14]The acceptance is defined by the interval [10,400] mrad in the polar angle. The limit at 400 mrad is a bit larger than the actual detector acceptance ($\sim$ 330 mrad) to avoid losses of events due to the magnetic field.

The generation acceptance and reconstruction efficiency, and its product are summarised in table 5.21. They have been computed using MC simulation, and its uncertainty is smaller than 0.1%.

	$B^0_s \to \mu^+\mu^-$	$B^+ \to J/\psi K^+$	$B^0_s \to J/\psi\,\phi$	$B^0_{(s)} \to h^+h'^-$	$B^0 \to J/\psi K^{*0}$
$\epsilon_{\rm gen}$	17.5%	15.3%	16.0%	17.7%	14.8%
$\epsilon_{\rm reco/gen}$	58.3%	38.6%	25.0%	44.5%	24.8%
total efficiency					
$\epsilon_{\rm reco/gen} \times \epsilon_{\rm gen}$	10.2%	5.9%	4.0%	7.9%	3.7%

Table 5.21: Generator level efficiency $\epsilon_{\rm gen}$ and reconstruction efficiencies $\epsilon_{\rm reco/gen}$ for signal and control channels. The muonID efficiency (for $B^0_{(s)} \to h^+h'^-$ acceptance) is taken into account in $\epsilon_{\rm reco/gen}$. The uncertainty on the quoted numbers is below 0.1%.

When normalising to $B^0_{(s)} \to h^+h'^-$, the two hadrons are required to be within the muon detector acceptance to minimise the differences between the signal and this control channel. The uncertainties associated to the muon system acceptance will be treated together with the muonID efficiency and described later in this subsection.

The ratio of the reconstruction efficiencies depends on the reconstruction efficiency of an extra track (the kaon) when the normalisation is done with the $B^+ \to J/\psi K^+$ channel, of two extra tracks when $B^0_s \to J/\psi\phi$ is used and on the different phase space of the muons in the final state for signal and $J/\Psi \to \mu^+\mu^-$.

The track reconstruction efficiency is determined from the simulation and corrected using the tracking efficiency map as provided by the LHCb tracking group [177]. More details on these corrections are given in [173].

5.9.3.2 Validation of the ratio of reconstruction efficiencies using data

The signal $B^0_{(s)} \to \mu^+\mu^-$ decay and the control channel $B^+ \to J/\psi K^+$ differ by one track in the final state. The acceptance and reconstruction efficiency for this extra-track can be probed by using the ratio of events of $B^+ \to J/\psi K^+$ with respect $B^0 \to J/\psi K^{*0}$, which contains four tracks in the final state. The ratio of reconstruction efficiencies is expected to follow:

$$\frac{\epsilon^{\rm REC}(B^+ \to J/\psi K^+)}{\epsilon^{\rm REC}(B^0_s \to \mu^+\mu^-)} \simeq \frac{\epsilon^{\rm REC}(B^0 \to J/\psi K^{*0})}{\epsilon^{\rm REC}(B^+ \to J/\psi K^+)}\,. \tag{5.18}$$

The ratio in equation 5.18 has been computed both in data and MC. The selection cuts for $B^0 \to J/\psi K^{*0}$ and $B^+ \to J/\psi K^+$ selections can be found in table 5.22. A simulated $b\bar{b} \to \mu^+\mu^- X$ sample has also been used, selecting both channels by applying the same selection than in data. Figure 5.43 shows the fit to the invariant mass of the $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$ distributions both in data and MC. The simulated sample has been prepared as described in section 5.9.4.1.

	Cut
$B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$	VDS(J/ψ) >25
	PIDk > 0
$B^0 \to J/\psi K^{*0}$	IPS(K^*) >5
	IPS($K^\pm$) >2
	IPS($\pi^\pm$) >2
	Δm (K^*) < 40 MeV/c^2
	vertex $\chi^2(K^*) < 9$

Table 5.22: Selection of $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$ used to obtain the ratio of reconstruction efficiencies.

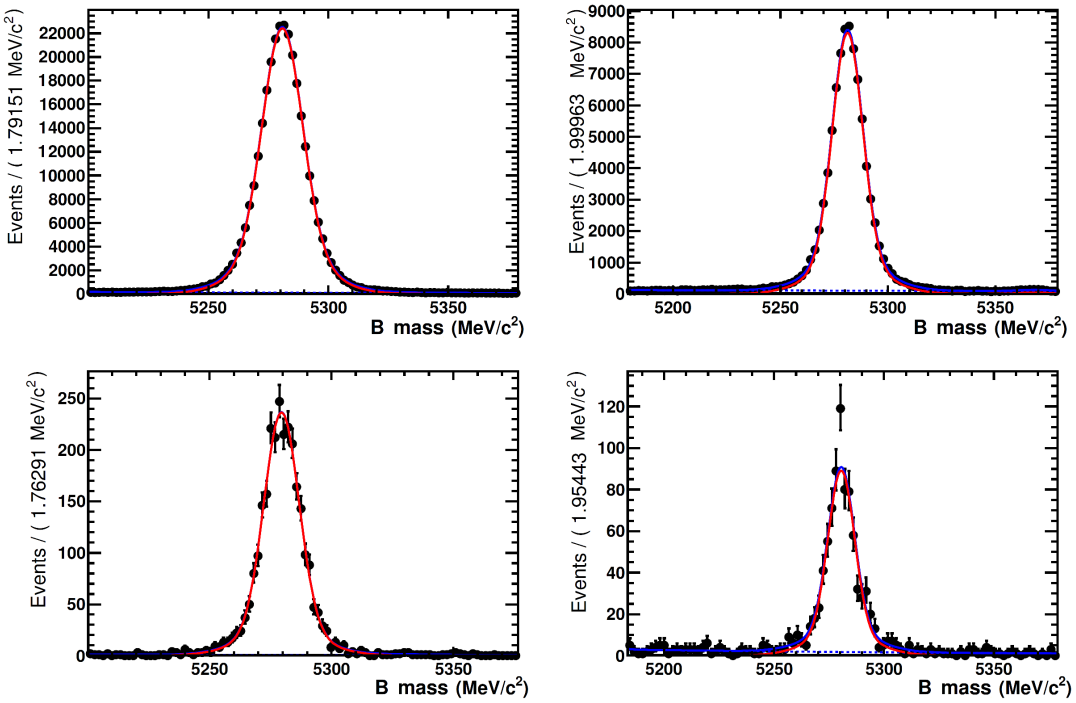

Figure 5.43: Invariant mass distributions of $B^+ \to J/\psi K^+$ (left) and $B^0 \to J/\psi K^{*0}$ (right) candidates after the selection described in the text. Top, 2011 data. Bottom MC10.

The following number of events are obtained:

$$N_{data}(B^+ \to J/\psi K^+) = 291\,632 \pm 560 \quad N_{MC}(B^+ \to J/\psi K^+) = 2653 \pm 60$$
$$N_{data}(B^0 \to J/\psi K^{*0}) = 84184 \pm 297 \quad N_{MC}(B^0 \to J/\psi K^{*0}) = 802 \pm 29$$

From these values, the ratio data/MC of the ratio of the reconstruction efficiencies is estimated to be:

$$\frac{\epsilon^{\rm REC,data}_{B^+\to J/\psi K^+} / \epsilon^{\rm REC,data}_{B^0\to J/\psi K^{*0}}}{\epsilon^{\rm REC,MC}_{B^+\to J/\psi K^+} / \epsilon^{\rm REC,MC}_{B^0\to J/\psi K^{*0}}} = 1.032 \pm 0.04 \pm 0.06 \tag{5.19}$$

where the first error is due to the statistics (dominated by the number of MC events) and the second is the uncertainty associated with the $\mathcal{B}$ of $B^0 \to J/\psi K^{*0}$ and $B^+ \to J/\psi K^+$. This estimation from the data validates the use of the reconstruction efficiency from simulation.

5.9.3.3 Muon detector acceptance and muonID efficiency

In section 5.5 the muon detector acceptance (acc_μ) and the muonID efficiency (ϵ_μ) have been determined as a function of p and p_T using different control samples.

The acceptance×efficiency 2D map in p and p_T bins obtained both in data and in simulated events are then folded into the p, p_T spectrum of the muons from reconstructed and selected $B^0_s \to \mu^+\mu^-$, $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$ Monte Carlo events and the average values $< \epsilon_\mu \times acc_\mu >_{\rm data}$ and $< \epsilon_\mu \times acc_\mu >_{\rm MC}$ are obtained. The ratio $C_{\mu ID}(B^0_s \to \mu^+\mu^-, B^+ \to J/\psi K^+, B^0_s \to J/\psi\phi) =< \epsilon_\mu \times acc_\mu >_{\rm data} / < \epsilon_\mu \times acc_\mu >_{\rm MC}$ is then used to correct the muonID efficiency evaluated on Monte Carlo for a given channel. For the $B^0_{(s)} \to h^+h'^-$ sample, the selection efficiency obtained in Monte Carlo is corrected only for the ratio $C_{\mu ID}(B^0_{(s)} \to$

$h^+h'^-) =< acc_\mu >_{\rm data} / < acc_\mu >_{\rm MC}$. More details in the determination corrections and on the systematic errors assigned can be found in [173].

The ratios of these corrections are used in the computation of the normalisation factors. In the case of the ratio between the $B^+ \to J/\psi K^+$, $B^0_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ and $B^0_{(s)} \to \mu^+\mu^-$ the following corrections are obtained:

$$\frac{C_{\mu ID}(B \to J/\psi X)}{C_{\mu ID}(B^0_{(s)} \to \mu^+\mu^-)} = 1.0108 \pm 0.0011_{stat} \pm 0.0004_{sample} \pm 0.0073_{trigger}, \tag{5.20}$$

where the differences between data and MC largely cancel in the ratio between $B^0_{(s)} \to \mu^+\mu^-$ and $J/\psi \to \mu^+\mu^-$.

In the case of the ratio between $B^0_{(s)} \to h^+h'^-$ and $B^0_{(s)} \to \mu^+\mu^-$ (it must be reminded that the muon detector acceptance is also required for $B^0_{(s)} \to h^+h'^-$), the same technique yields:

$$\frac{C_{\mu ID}(B^0_{(s)} \to h^+h'^-)}{C_{\mu ID}(B^0_{(s)} \to \mu^+\mu^-)} = 0.9698 \pm 0.0006_{stat} \pm 0.0042_{sample} \pm 0.0032_{trigger}, \tag{5.21}$$

which is entirely dominated by the remaining data-MC differences in the muonID efficiencies for $B^0_{(s)} \to \mu^+\mu^-$ (section 5.5).

Finally, the interaction of the hadrons with the detector material has also been taken into account in the case of normalisation to $B^0_{(s)} \to h^+h'^-$. From [168], this effect introduces an extra uncertainty in the tracking efficiency reconstruction of 1% per track.

5.9.4 Ratio of selection efficiencies

The selection criteria for signal and normalisation channels were described in section 5.4.

As in the case of reconstruction, the selection efficiencies are determined using MC simulation and then cross-checked with data. The main difference between the MC and the data is in the IP distribution. In order to minimise this difference, a technique developed by the tracking group to smear on the MC sample the parameters of the tracks at the origin has been applied, thus reproducing the observed IP resolution in data. The selection efficiencies are then determined from the smeared MC.

As the uncertainties cancel in the ratio of efficiencies, no systematic uncertainties in the ratio of selection efficiencies have been assigned for the cuts that are identical between the signal and the control channel. In opposition, a systematic error on the ratio of selection efficiencies has been estimated on the cuts that are different between the signal and the control channels.

5.9.4.1 Track parameter smearing

The tracking group has proposed a method [178] to adjust the MC precision of the first state of a track to the one observed in data. The width of the IP distribution of minimum bias tracks is parameterised as a function of p_T and ϕ (the angle of the track in the transverse plane with respect the x direction) for both data and MC. The difference of both distributions is then used as width for a Gaussian smearing of the MC. This method has been applied to the MC tracks of the B decay products, and all the quantities of the selection have been recomputed.

5.9.4.2 Selection efficiencies from simulation

The selection efficiencies have been computed using the MC in three different scenarios: the original (unsmeared) MC, the smeared MC and a MC where the smearing is overdone by 50% to estimate possible residual effects. Notice that the selection efficiency is determined on events where all signal tracks have been reconstructed.

The biggest difference in the selection efficiencies for the $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$ normalisation channels comes from additional IPχ^2 requirements on the extra tracks in the

normalisation channel, and, to a smaller extent, from the different kinematics of the muons from the J/ψ and the B_s^0.

For the $B^0_{(s)} \to h^+h'^-$ normalisation channel, the selection efficiency is higher than for the signal as the tight (60 MeV) mass window is not applied to the $B^0_{(s)} \to h^+h'^-$ channel.

To validate the smearing procedure, the selection variables can be compared in data and MC. This is done in figure 5.44, using the channel $B^+ \to J/\psi K^+$, chosen because of being almost background free and having large statistics.

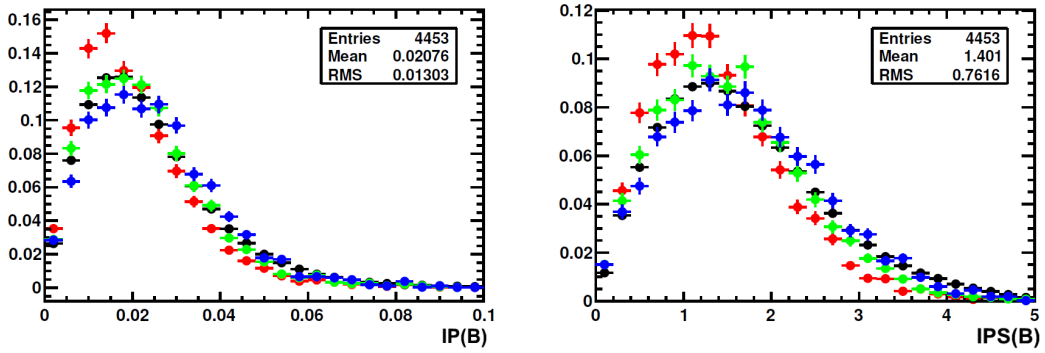

Figure 5.44: Comparison of variables in MC after different smearings with data, using $B^+ \to J/\psi K^+$. In all figures, black : data, red: unsmeared MC, green: smeared MC, blue: oversmeared MC. Left, IP(B). Right IPS(B). It can seen that the three MC samples cover the whole range of the data distribution.

5.9.4.3 PID efficiency determination

The RICH PID efficiency for $B_s^0 \to J/\psi\phi$ is not included in table 5.23 as it is known that the simulation does not describe the PID efficiency in data correctly. Instead, it is measured directly on data using two different tight selection criteria (additional cuts: $\tau > 1$ ps, B vertex $\chi^2 < 25$ or all track IPS > 9, B VDS >25). It has been verified on MC that the PID efficiency after the tight selection is within 1% identical to the original RICH PID.

The PID efficiency of $DLL(K-\pi) > 0$ for both kaons has been determined to be

$$\epsilon = (91.4 \pm 0.9_{\text{stat}} \pm 1.0_{\text{syst}})\% \tag{5.22}$$

efficient and 1% additional systematic uncertainty is added to account for selection dependent PID effects.

channel	normal	smeared	50% over smeared
$B_s^0 \to \mu^+\mu^-$	49.8%	47.2%	44.0%
$B^0 \to K^+\pi^-$	55.1%	52.5%	48.8%
$B^+ \to J/\psi K^+$	41.7%	39.5%	36.3%
$B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$	32.3%	30.4%	28.1%

Table 5.23: Selection efficiencies for signal and control channels. The efficiencies include the efficiency of the fiducial cuts as well as the BDTS cut efficiencies.

5.9.4.4 Ratio of selection efficiencies

The total selection efficiency for the signal and control channels is given in table 5.23. The ratio of selection efficiencies for the three normalisation channels is determined from the smeared MC, the values being:

$$\epsilon^{\text{SEL|REC}}_{B^+\to J/\psi K^+} / \epsilon^{\text{SEL|REC}}_{B^0_{(s)}\to\mu^+\mu^-} = 0.836 \pm 0.01 \tag{5.23}$$

$$\epsilon^{\text{SEL|REC}}_{B^0_s\to J/\psi\,(\mu^+\mu^-)\phi(K^+K^-)} / \epsilon^{\text{SEL|REC}}_{B^0_{(s)}\to\mu^+\mu^-} = 0.589 \pm 0.02 \tag{5.24}$$

$$\epsilon^{\text{SEL|REC}}_{B^0\to K^+\pi^-} / \epsilon^{\text{SEL|REC}}_{B^0_{(s)}\to\mu^+\mu^-} = 1.11 \pm 0.01 \tag{5.25}$$

where the errors are from MC statistics.

The absolute selection efficiencies between the unsmeared MC and the over-smeared sample vary by $5-7\%$, depending on the channel. However, the ratio between signal and all three normalisation channels stays constant within 1%, since the efficiency change in the signal is canceled by the corresponding change in the normalisation channel. For more details on the systematics on the ratio of selection efficiencies see [173].

5.9.5 Ratio of trigger efficiencies

The trigger efficiencies are estimated using the TISTOS method [179]. For the control samples, this method can be directly applied on data, while for the $B^0_{(s)} \to \mu^+\mu^-$ it has to be done in two steps. First, the efficiency with the TISTOS method on $J/\psi \to \mu\mu$ events selected by a detached J/ψ selection is determined in bins of the p_T and IP. Then, this efficiency is reweighted using the $B^0_s \to \mu^+\mu^-$ spectra taken from the MC. Only the muon trigger lines are used in this process.

The trigger lines that select the $B^0_{(s)} \to \mu^+\mu^-$ channel are the same which select the normalisation channels containing muons in the final state, $B^+ \to J/\psi K^+$ and $B^0_s \to J/\psi\phi$. In this case the ratio of trigger efficiencies is not expected to be very different from 1, being the differences manly due to the different phase space covered by the muons in the final state.

On the contrary, large corrections are expected in the case of the normalisation with a fully hadronic channel as the $B^0 \to K^+\pi^-$. But, as it will be described in section 5.9.6.3, some factors cancel in the ratio and finally the normalisation is done to $B^0_{(s)} \to h^+h'^-$ TIS events, so that what enters in the normalisation is the trigger TIS efficiency.

5.9.5.1 Estimation of the trigger efficiencies

As already said in this thesis, LHCb events can be classified in two categories [179]: TIS and TOS. TIS refers to events which would have also been triggered without the signal under study. TOS, to those in which the signal under study is sufficient to trigger the event. Note that an event can also be TIS and TOS simultaneously (TIS&TOS), or neither TOS nor TIS. The LHCb trigger system records all the information needed for such a classification.

The overall trigger efficiency on such events can be expressed as [89]:

$$\epsilon^{\text{TRIG}} = \frac{N^{\text{TRIG}}}{N^{\text{SEL}}} = \frac{N^{\text{TIS}}}{N^{\text{SEL}}}\frac{N^{\text{TRIG}}}{N^{\text{TIS}}} = \epsilon^{\text{TIS}}\,\frac{N^{\text{TRIG}}}{N^{\text{TIS}}} \tag{5.26}$$

Both N^{TRIG} and N^{TIS} are observable quantities, ϵ^{TIS} is the efficiency to trigger without any information from the signal.

As the TIS events are mainly triggered by the decay products of the other b quark in the event, the TIS efficiency ϵ^{TIS} can slightly depend on the p_T of the B signal via the $b\bar{b}$ correlation. However, as the signal and the normalisation channels were selected in the same way (see section 5.4), the ϵ^{TIS} is independent of the considered decay and cancels in the ratio.

The most precise measurement of $\epsilon^{\rm TIS}$ can be done in the $B^+ \to J/\psi K^+$ channel, and it is found to be

$$\epsilon^{\rm TIS}({\rm L0} \times {\rm HLT1}) = (5.13 \pm 0.9_{stat} \pm 0.4_{syst})\% \,, \tag{5.27}$$

the first error being statistical (dominated by the number of TIS events) and the second systematic. The systematic error has been evaluated in MC from the difference of the trigger efficiencies between the true value and the one obtained from the TISTOS method.

Since the requirement of HLT2 TIS would drastically reduce the samples, $\epsilon^{\rm TIS}({\rm L0} \times {\rm HLT1})$ is considered, using when needed the HLT2 efficiencies from the MC and assigning a systematic uncertainty to it.

5.9.5.2 Ratio of trigger efficiency for $B^0_{(s)} \to \mu^+\mu^-$ and $B \to J/\psi X$ channels

The trigger efficiency for the $B \to J/\psi X$ channels estimated using the TISTOS method is:

$$\epsilon^{\rm TRIG|SEL}_{J/\psi} \quad = \quad (87.2 \pm 0.4_{\rm stat} \pm 3.8_{\rm syst})\% \,, \tag{5.28}$$

The systematic error is the combination of two: one associated to the TISTOS method precision obtained from MC and the second one due to the 2.3% of $B \to J/\psi X$ events that are triggered not using the muon triggers.

As explained, and in order to estimate the trigger efficiency for the signal, an efficiency map has been computed (see figure 5.45) as a function of the largest p_T and largest IP of the muons from the $B^+ \to J/\psi K^+$ detached selection.

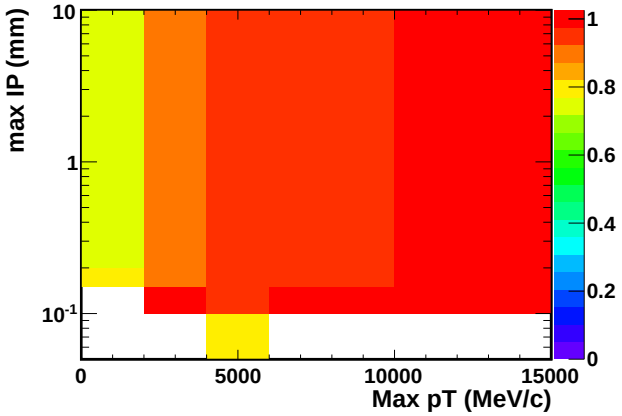

Figure 5.45: Trigger efficiency map as a function of the max p_T and $\max_{\rm IP}$ of the muons from $B \to J/\psi X$ events.

Several variables have been tested for the mapping, with the ones chosen providing the smallest difference between the TISTOS method and the true efficiency in MC. This efficiency map is then applied to the muon spectrum of the $B^0_s \to \mu^+\mu^-$ MC sample. The estimated trigger efficiency for $B^0_s \to \mu^+\mu^-$ events is:

$$\epsilon^{\rm TRIG|SEL}_{B^0_s \to \mu^+\mu^-} = (91.4 \pm 0.4_{\rm stat} \pm 3.9_{\rm syst})\%, \tag{5.29}$$

Again, the systematic error is the combination of two: one associated to the TISTOS method applied to $B^0_s \to \mu^+\mu^-$ and the second one due to the 2.3% of $B^0_s \to \mu^+\mu^-$ events that are triggered not using the muon triggers.

The ratio of trigger efficiencies between the signal and the $B^+ \to J/\psi K^+$ and the $B^0_s \to J/\psi\phi$

normalisation channels is then computed as the ratio of the two efficiencies:

$$\frac{\epsilon^{\rm TRIG|SEL}_{B^+\to J/\psi K^+}}{\epsilon^{TRIG|SEL}_{\rm B^0_s\to\mu^+\mu^-}} = \frac{\epsilon^{\rm TRIG|SEL}_{B^0_s\to J/\psi\,(\mu^+\mu^-)\phi(K^+K^-)}}{\epsilon^{TRIG|SEL}_{\rm B^0_s\to\mu^+\mu^-}} = (95.4 \pm 0.4_{\rm stat} \pm 1_{\rm syst})\%, \tag{5.30}$$

where the systematic error has been evaluated from MC. This error is small because the bias due to the TISTOS method is the same for the two channels and cancels in the ratio.

5.9.5.3 Global event cuts

The effect of GECs has been evaluated repeating the efficiency evaluation using only the TIS L0DiMuon, as it is the only L0 line with a loose SPD multiplicity cut (see section 3.2.3.1). The absolute efficiency values for J/ψ and $B^0_s \to \mu^+\mu^-$ go down by about 4% but the ratio remains constant within 0.1%. It can hence be concluded that the GECs do not introduce any additional uncertainty.

5.9.6 Number of candidates

5.9.6.1 Normalisation to $B^+ \to J/\psi K^+$

Figure 5.46 shows the invariant mass distribution of the events passing the $B^+ \to J/\psi K^+$ selection and the fit to the signal and background. The signal distribution is modeled with a double Gaussian function, while the background is modeled with two functions: an exponential for the combinatorial background, and a Gaussian on the right for the physical background. This one is associated to $B^+ \to J/\Psi\pi^+$ events. The number of signal events after the selection and a BDTS cut of 0.05 is: $N(B^+ \to J/\psi K^+) = 340\,129 \pm 640$. A systematic error of 1.3% has been assigned due to the differences between the number of candidates obtained with the fit and those obtained after subtracting the background. This number of signal events contains 0.1% duplicated candidates.

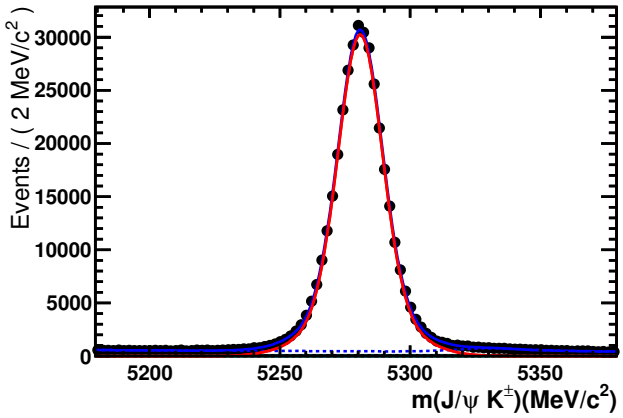

Figure 5.46: Invariant mass distribution of the $B^+ \to J/\psi K^+$ candidates after the stripping selection.

5.9.6.2 Normalisation to $B^0_s \to J/\psi\,(\mu^+\mu^-)\phi(K^+K^-)$

Figure 5.47 shows the invariant mass distribution of the events passing the $B^0_s \to J/\psi\,(\mu^+\mu^-)\phi(K^+K^-)$ selection and the fit to the signal and background. The signal distribution is modeled with a double Gaussian function, while the combinatorial background is modeled

with an exponential. The number of signal events after the selection and a BDTS cut of 0.05 is: $N(B^+ \to J/\psi K^+) = 19\,035 \pm 139$. The fit result has been compared to the number of candidates obtained after background subtraction and a 0.4% systematic uncertainty is added to account for the differences. The fraction of duplicated candidates in this signal is 0.2%.

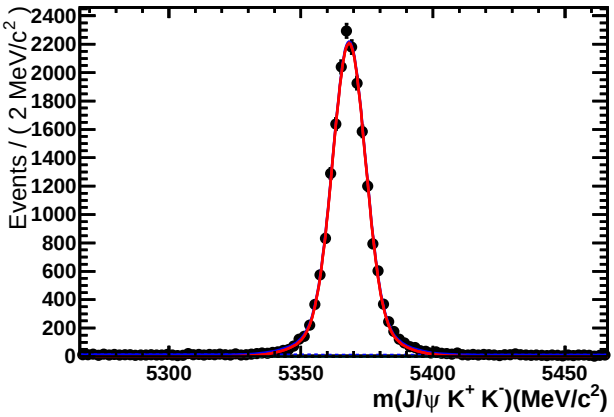

Figure 5.47: Invariant mass distribution of the $B_s^0 \to J/\psi\phi$ candidates after the selection.

5.9.6.3 Normalisation to $B^0 \to K^+\pi^-$ or $B^0_{(s)} \to h^+h'^-$ TIS

The $\alpha_{B^0\to K^+\pi^-}$ factor can be written in the following way:

$$\alpha_{B^0\to K^+\pi^-} = \mathcal{B}(B^0 \to K^+\pi^-) \times \frac{f_d}{f_s}\frac{\epsilon_{hh}^{REC}}{\epsilon_s^{REC}}\frac{\epsilon_{hh}^{SEL/REC}}{\epsilon_s^{SEL/REC}}\frac{\epsilon_{hh}^{TRIG/SEL}}{\epsilon_s^{TRIG/SEL}}\frac{1}{f_{B^0\to K^+\pi^-}N_{hh}} \tag{5.31}$$

where N_{hh} is the number of $B^0_{(s)} \to h^+h'^-$ events and $f_{B^0\to K^+\pi^-}$ is the fraction of the $B^0 \to K^+\pi^-$ in the inclusive sample. The reconstruction, selection, and trigger efficiencies are assumed to be the same for all the $B^0_{(s)} \to h^+h'^-$ modes. This fraction was measured in [168] to be: $f_{B^0\to K^+\pi^-} = 0.605 \pm 0.027$.

The $\epsilon_{hh}^{TRIG/SEL}$ can be written as:

$$\epsilon_{hh}^{TRIG/SEL} = \epsilon_{hh}^{TIS/SEL}\frac{N_{hh}}{N_{hh}^{TIS}} \tag{5.32}$$

Therefore, replacing equation 5.32 in 5.31, N_{hh} cancels out. This means that the normalisation to a given channel, when the trigger efficiency is measured in the same channel, becomes equivalent to a normalisation to TIS events:

$$\alpha = \mathcal{B}(B^0 \to K^+\pi^-) \times \frac{f_N}{f_{sig}}\frac{\epsilon_{hh}^{REC}}{\epsilon_s^{REC}}\frac{\epsilon_{hh}^{SEL/REC}}{\epsilon_s^{SEL/REC}}\frac{\epsilon_{hh}^{TIS/SEL}}{\epsilon_s^{TRIG/SEL}}\frac{1}{f_{B^0\to K^+\pi^-}N_{hh}^{TIS}} \tag{5.33}$$

ϵ^{TIS}(L0 × HLT1) was computed in section 5.9.5.1. From the MC simulation, the HLT2 efficiency on L0 and HLT1 $B^0_{(s)} \to h^+h'^-$ TIS events is estimated to be $\epsilon^{HLT2} = 83.6$ %, so that: $\epsilon_{hh}^{TIS/SEL} = (4.3 \pm 0.3)$ %. With all this, the ratio of trigger efficiencies becomes:

$$\frac{\epsilon_{hh}^{TIS/SEL}}{\epsilon_{B_s^0\to\mu^+\mu^-}^{TRIG|SEL}} = (4.7 \pm 0.3)\% \tag{5.34}$$

It should be also reminded that the total number of $B^0_{(s)} \to h^+h'^-$ has been computed when calibrating the BDT, and found to be: $N_{hh} = 16734 \pm 757$.

5.9.7 Normalisation factor

The normalisation factor α_{cal} for the different control channels and the factors that enters in its calculation, that have been presented along this section, are listed in table 5.24. To obtain α^{cal} $f_d/f_s = 3.75 \pm 0.29$ [176] has been used. The only difference between $\alpha^{cal}_{B^0\to\mu^+\mu^-}$ and $\alpha^{cal}_{B^0_s\to\mu^+\mu^-}$ is the use (or not) of f_d/f_s depending on the normalisation channel.

	$\mathcal{B}$	$\frac{\epsilon^{\rm REC}_{\rm cal}\epsilon^{\rm SEL\|REC}_{\rm cal}}{\epsilon^{\rm REC}_{\rm sig}\epsilon^{\rm SEL\|REC}_{\rm sig}}$	$\frac{\epsilon^{\rm TRIG\|SEL}_{\rm cal}}{\epsilon^{\rm TRIG\|SEL}_{\rm sig}}$	N_{cal}	$\alpha^{cal}_{B_d\to\mu^+\mu^-}$	$\alpha^{cal}_{B_s\to\mu^+\mu^-}$
	$(\times 10^{-5})$				$(\times 10^{-11})$	$(\times 10^{-10})$
$B^+ \to J/\psi K^+$	6.01 ± 0.21	0.502 ± 0.013	0.954 ± 0.022	$340\,129 \pm 4468$	8.464 ± 0.433	3.170 ± 0.297
$B^0_s \to J/\psi\phi$	3.4 ± 0.9	0.245 ± 0.011	0.954 ± 0.022	$19\,035 \pm 158$	11.13 ± 3.124	4.169 ± 1.123
$B^0 \to K^+\pi^-$	1.94 ± 0.06	0.857 ± 0.028	0.0469 ± 0.0034	$10\,124 \pm 916$	7.709 ± 0.957	2.887 ± 0.424

Table 5.24: Summary of the factors and their uncertainty entering in the normalisation for the three channels considered.

A weighted average taking the tracking and trigger uncertainties to be correlated between the two J/ψ normalisation channels, and the uncertainty on f_d/f_s to be correlated between the $B^+ \to J/\psi K^+$ and $B^0 \to K^+\pi^-$, gives:

$$\alpha_{B^0_s\to\mu^+\mu^-} = (3.19 \pm 0.28) \times 10^{-10}\,, \tag{5.35}$$

$$\alpha_{B^0\to\mu^+\mu^-} = (8.38 \pm 0.39) \times 10^{-11}\,, \tag{5.36}$$

which are the normalisation factors used in the computation of the limits. Using these factors and the SM predictions for the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ (see section 2.1.3) $\sim$1 $B^0 \to \mu^+\mu^-$ and $\sim$11 $B^0_s \to \mu^+\mu^-$ would be expected in total in the analysed sample.

5.10 Results: upper limits

The expected number of combinatorial background events, peaking background events and signal events assuming the SM branching ratios, together with the number of observed events, are shown in table 5.25 for $B^0 \to \mu^+\mu^-$ analysis and in table 5.26 for $B^0_s \to \mu^+\mu^-$ analysis for each of the 72 bins (8 bins in BDT and 9 bins in invariant mass) that define the signal regions. The uncertainties in the signal and background likelihoods and normalisation factors are used to compute the uncertainties in the background and signal predictions.

The distribution of events in the invariant mass vs. BDT plane can be seen in figure 5.48. The distribution of the invariant mass in the eight BDT bins is shown in figure 5.49 for $B^0 \to \mu^+\mu^-$ candidates and in figure 5.50 for $B^0_s \to \mu^+\mu^-$ candidates. The same distributions obtained by integrating the five most significant BDT bins and with equally spaced mass bins are shown in figure 5.51. In the limit computation the cross-feed of $B^0_s \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) events in the B^0 (B^0_s) mass window has been taken into account assuming the SM rates.

The compatibility of the distribution of events in the invariant mass vs. BDT plane with a given branching ratio hypothesis is evaluated using the $\rm CL_s$ method (see [152, 153] and section 5.7.2.1). As explained, this method provides two estimators: $\rm CL_s$ is a measure of the compatibility of the observed distribution with the signal hypothesis, while $\rm CL_b$ is a measure of the compatibility with the background-only hypothesis.

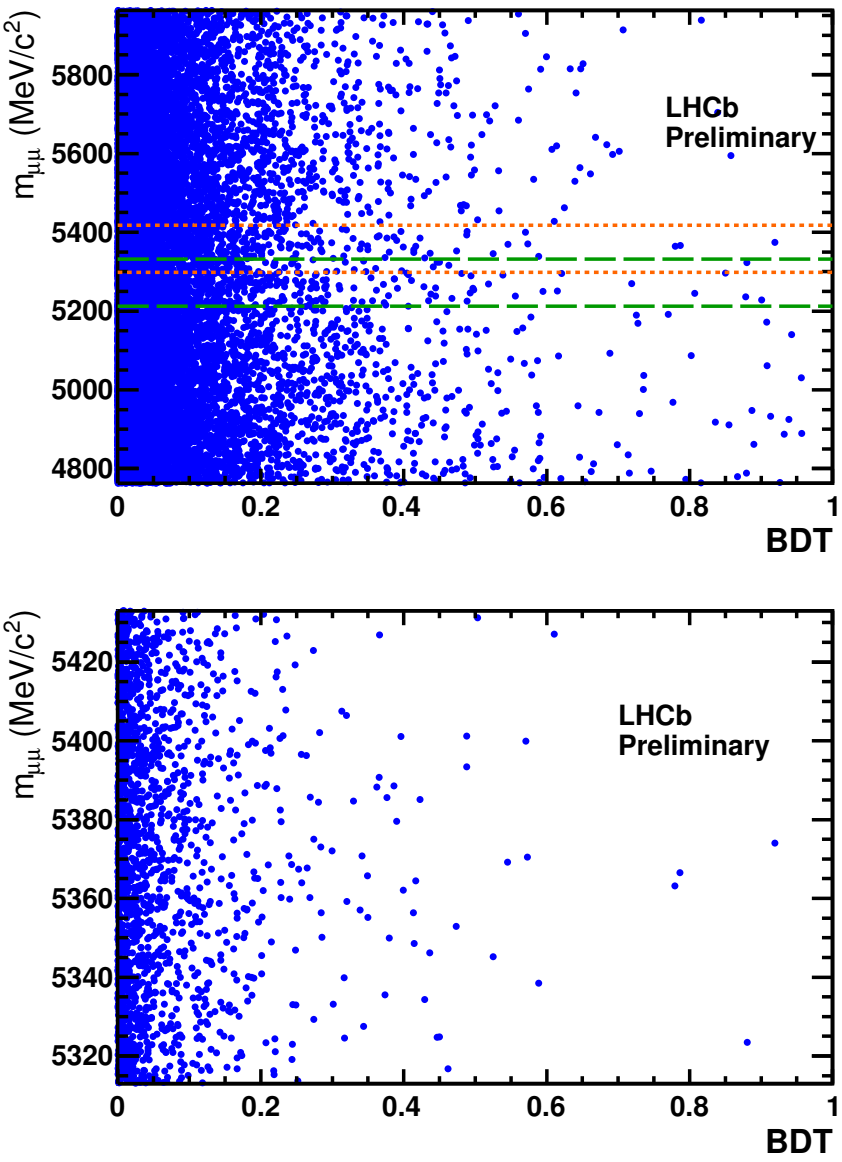

Figure 5.48: Unblinded data: 2D plot of mass versus BDT. Top: green long-dashed (orange short-dashed) lines indicate the $\pm 60\,\mathrm{MeV}/c^2$ search window around the B^0 (B_s^0). Bottom: zoom in the B_s^0 search window.

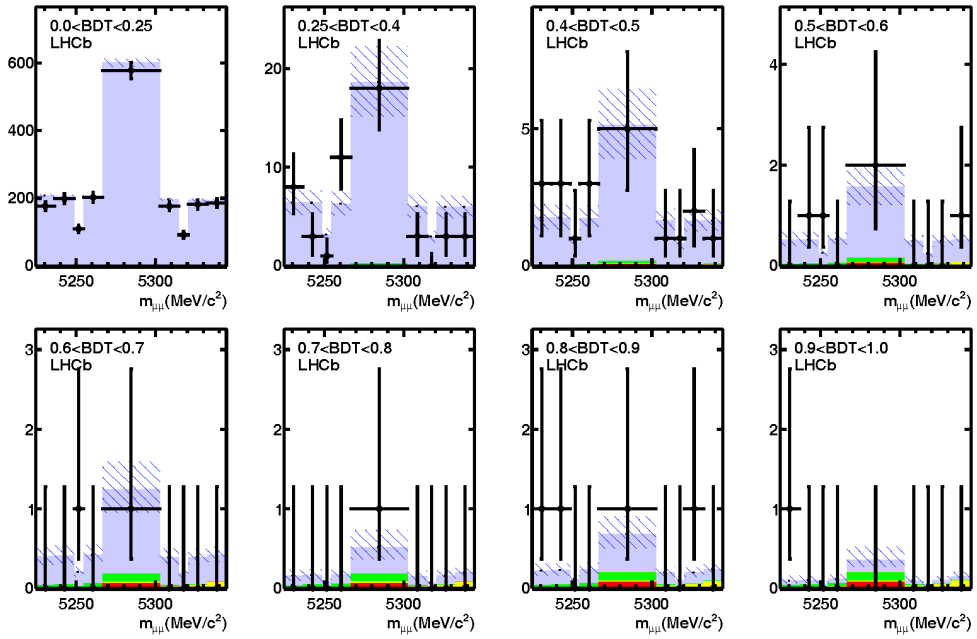

Figure 5.49: $B^0 \to \mu^+\mu^-$: distribution of selected dimuon events in the invariant mass plane for the eight BDT bins. The black dots are data, the light blue histogram shows the contribution of the combinatorial background, the green histogram shows the contribution of the $B^0_{(s)} \to h^+h'^-$ background and the yellow histogram shows the cross-feed of $B^0_s \to \mu^+\mu^-$ events in the B^0 mass window assuming the SM rate. The hatched area depicts the uncertainty on the sum of the expected contributions.

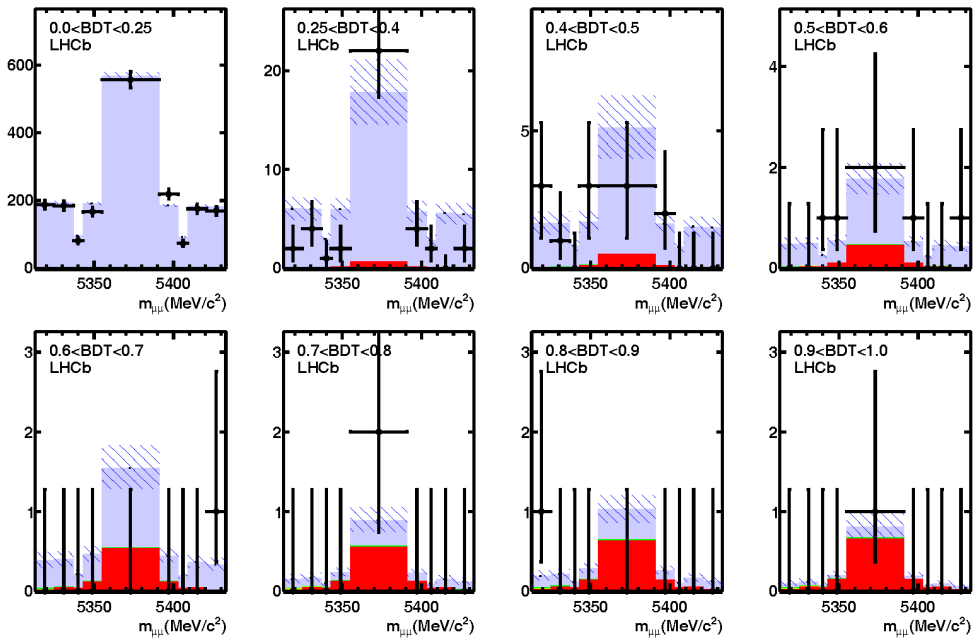

Figure 5.50: $B_s^0 \to \mu^+\mu^-$: distribution of selected dimuon events in the invariant mass plane for the eight BDT bins. The black dots are data, the light blue histogram shows the contribution of the combinatorial background, the green histogram shows the contribution of the $B^0_{(s)} \to h^+h'^-$ background and the red filled histogram the contribution of $B_s^0 \to \mu^+\mu^-$ signal events according to the SM rate. The hatched area depicts the uncertainty on the sum of the expected contributions.

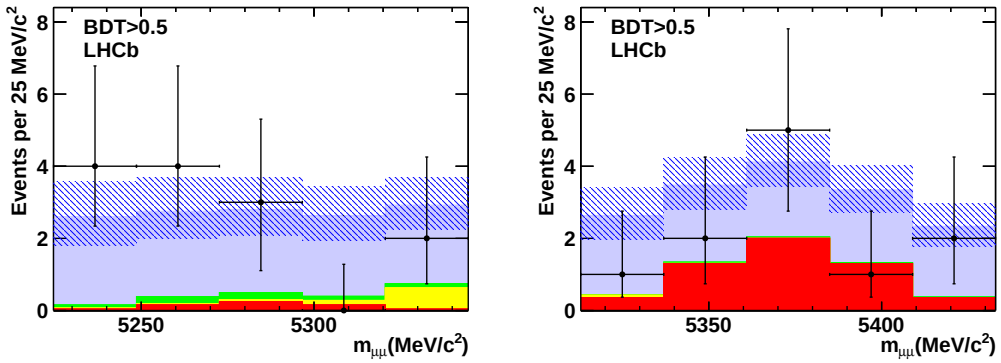

Figure 5.51: $B^0 \to \mu^+\mu^-$ (left) and $B_s^0 \to \mu^+\mu^-$ (right) distribution of selected dimuon events in the invariant mass plane for events with BDT>0.5. The black dots are data, the light blue histogram shows the contribution of the combinatorial background, the green histogram shows the contribution of the $B^0_{(s)} \to h^+h'^-$ background and the red filled histogram the contribution of signal events according to the SM rate. The hatched area depicts the uncertainty on the sum of the expected contributions; differently from what used for the limit evaluation, equally spaced mass bins are shown.

For the $B^0 \to \mu^+\mu^-$ decay, the distribution of expected (dashed black lines) and observed CL_s (dotted blue line) is shown in figure 5.52 as a function of the assumed branching ratio. The green shaded areas cover the region of $\pm 1\sigma$ of compatible observations. The limits obtained at 90% and 95% CL are listed in table 5.27, together with the observed CL_b. For B^0, the comparison between data and expected background results in a p-value (1-CL_b) of 60%.

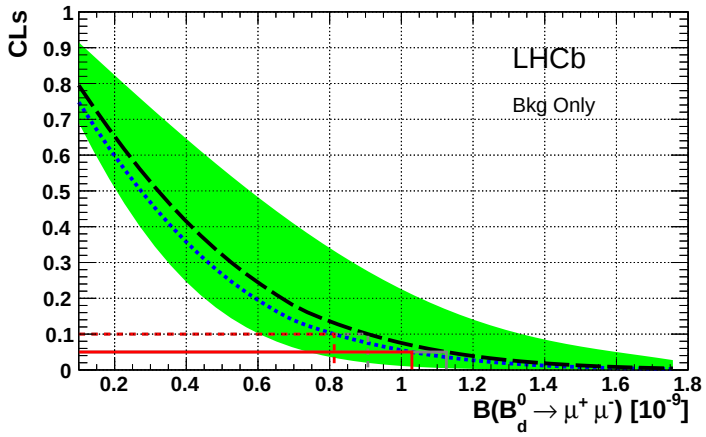

Figure 5.52: $B^0 \to \mu^+\mu^-$: expected CL_s (dashed black line) under the hypothesis to observe background-only, with green area covering the region of $\pm 1\sigma$ of compatible observations; the observed CL_s is given by the blue dotted line; the expected (observed) upper limits at 90% and 95% C.L. are also shown as dashed and solid grey (red) lines.

For the $B^0_s \to \mu^+\mu^-$ decay, the distribution of expected CL_s values is shown as dashed (black) lines in figure 5.53 under the hypothesis to observe background-only (left) or a combination of background-plus-SM events (right). Again, the green shaded areas cover the region of $\pm 1\sigma$ of compatible observations, and the observed CL_s as a function of the assumed branching ratio is shown as a dotted (blue) line. The expected and observed limits at 90 % and 95 % CL are listed in table 5.28, also with the observed CL_b. For the $B^0_s \to \mu^+\mu^-$ search window, the comparison of the distributions of observed events and expected background events results in a p-value of 18%. The probability that the observed events are compatible with the sum of expected background and signal events according to the SM rate is $\sim 17\%$.

As explained in the introduction of this chapter, around 12 $B^0_s \to \mu^+\mu^-$ triggered, reconstructed and selected events would be expected in the 2011 dataset in case the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ was the one predicted by the SM. So, even if no significant signal excess above the background has been found, some of the observed candidates are likely to be genuine $B^0_s \to \mu^+\mu^-$ decays. One particular event has been tagged as the favourite to be a SM candidate. Having a BDT of 0.9 and $m_{\mu\mu} = 5357$ MeV/c^2, this candidate is disfavoured to be neither $b\bar{b} \to \mu\mu X$ nor $B^0_{(s)} \to h^+h'^-$, since it belongs to one of the most sensitive bins, where almost no background and precisely around 1 SM candidates are expected (see table 5.26). Figure 5.54 shows an event display of the candidate, where the muons of the final state can be easily distinguished.

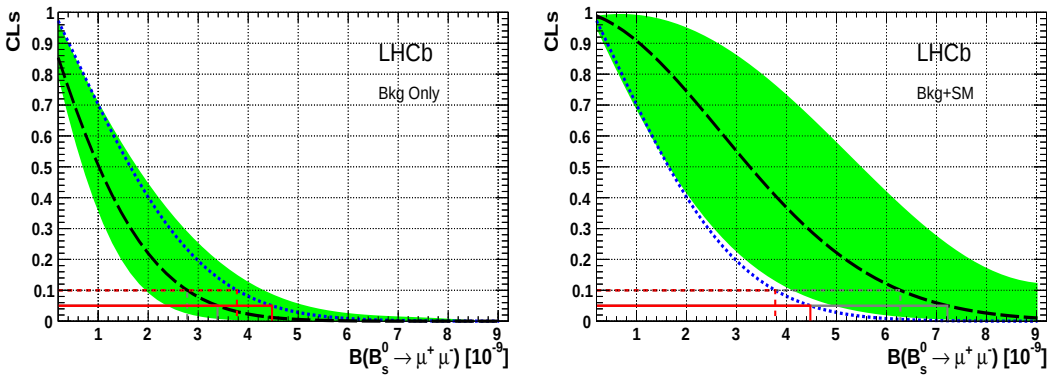

Figure 5.53: $B_s^0 \to \mu^+\mu^-$: expected CL_s (dashed black line) under the hypothesis to observe background-only (left) and a combination of background-plus-signal events according to the SM rate (right), with green area covering the region of $\pm 1\sigma$ of compatible observations; in both plots the observed CL_s is given by the blue dotted line; the expected (observed) upper limits at 90% and 95% C.L. are also shown as dashed and solid grey (red) lines.

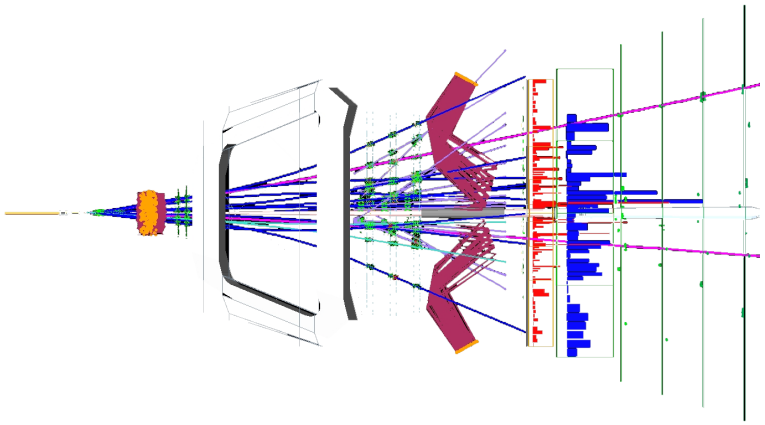

Figure 5.54: Event display of the best candidate for SM $B_s^0 \to \mu^+\mu^-$, built with Panoramix. Both muons of the final state can be easily distinguished (magenta lines).

Invariant mass [MeV/c^2]		BDT							
		0.0 – 0.25	0.25 – 0.4	0.4 – 0.5	0.5 – 0.6	0.6 – 0.7	0.7 – 0.8	0.8 – 0.9	0.9 – 1.0
5224 – 5236	Exp. comb. bkg	$207.6^{+4.5}_{-4.5}$	$6.4^{+1.3}_{-1.2}$	$1.74^{+0.47}_{-0.45}$	$0.49^{+0.14}_{-0.13}$	$0.37^{+0.13}_{-0.11}$	$0.121^{+0.074}_{-0.059}$	$0.187^{+0.087}_{-0.067}$	$0.063^{+0.061}_{-0.033}$
	Exp. peak. bkg	$0.084^{+0.043}_{-0.031}$	$0.042^{+0.017}_{-0.010}$	$0.0333^{+0.0131}_{-0.0079}$	$0.0297^{+0.0117}_{-0.0065}$	$0.0334^{+0.0130}_{-0.0073}$	$0.0315^{+0.0122}_{-0.0067}$	$0.0328^{+0.0128}_{-0.0070}$	$0.0317^{+0.0129}_{-0.0081}$
	Exp. cross-feed	$0.0064^{+0.0018}_{-0.0019}$	$0.00310^{+0.00056}_{-0.00050}$	$0.00246^{+0.00040}_{-0.00036}$	$0.00219^{+0.00030}_{-0.00026}$	$0.00253^{+0.00034}_{-0.00030}$	$0.00263^{+0.00033}_{-0.00029}$	$0.00300^{+0.00039}_{-0.00034}$	$0.00311^{+0.00060}_{-0.00055}$
	Exp. signal	$0.0055^{+0.0017}_{-0.0016}$	$0.00266^{+0.00054}_{-0.00047}$	$0.00211^{+0.00039}_{-0.00034}$	$0.00188^{+0.00030}_{-0.00026}$	$0.00218^{+0.00034}_{-0.00030}$	$0.00226^{+0.00034}_{-0.00030}$	$0.00257^{+0.00040}_{-0.00035}$	$0.00266^{+0.00057}_{-0.00050}$
	Observed	177	8	3	0	0	0	1	1
5236 – 5248	Exp. comb. bkg	$205.9^{+4.5}_{-4.5}$	$6.3^{+1.2}_{-1.2}$	$1.72^{+0.46}_{-0.44}$	$0.49^{+0.14}_{-0.13}$	$0.37^{+0.12}_{-0.11}$	$0.119^{+0.073}_{-0.058}$	$0.181^{+0.084}_{-0.065}$	$0.061^{+0.059}_{-0.032}$
	Exp. peak. bkg	$0.096^{+0.054}_{-0.036}$	$0.049^{+0.022}_{-0.013}$	$0.038^{+0.017}_{-0.010}$	$0.0346^{+0.0147}_{-0.0091}$	$0.039^{+0.016}_{-0.010}$	$0.0371^{+0.0149}_{-0.0100}$	$0.038^{+0.016}_{-0.010}$	$0.036^{+0.017}_{-0.010}$
	Exp. cross-feed	$0.0073^{+0.0021}_{-0.0021}$	$0.00351^{+0.00063}_{-0.00057}$	$0.00279^{+0.00045}_{-0.00040}$	$0.00248^{+0.00034}_{-0.00030}$	$0.00287^{+0.00038}_{-0.00034}$	$0.00298^{+0.00037}_{-0.00033}$	$0.00340^{+0.00044}_{-0.00039}$	$0.00353^{+0.00068}_{-0.00062}$
	Exp. signal	$0.0136^{+0.0040}_{-0.0040}$	$0.0065^{+0.0013}_{-0.0011}$	$0.00519^{+0.00092}_{-0.00082}$	$0.00463^{+0.00070}_{-0.00063}$	$0.00537^{+0.00079}_{-0.00072}$	$0.00556^{+0.00079}_{-0.00071}$	$0.00634^{+0.00093}_{-0.00083}$	$0.0066^{+0.0014}_{-0.0012}$
	Observed	199	3	3	1	0	0	1	0
5248 – 5254	Exp. comb. bkg	$102.3^{+2.2}_{-2.2}$	$3.15^{+0.61}_{-0.60}$	$0.85^{+0.23}_{-0.22}$	$0.241^{+0.068}_{-0.062}$	$0.182^{+0.061}_{-0.053}$	$0.059^{+0.036}_{-0.029}$	$0.088^{+0.041}_{-0.032}$	$0.030^{+0.029}_{-0.015}$
	Exp. peak. bkg	$0.051^{+0.030}_{-0.019}$	$0.0259^{+0.0122}_{-0.0074}$	$0.0204^{+0.0093}_{-0.0058}$	$0.0186^{+0.0080}_{-0.0052}$	$0.0209^{+0.0088}_{-0.0059}$	$0.0200^{+0.0080}_{-0.0057}$	$0.0206^{+0.0086}_{-0.0058}$	$0.0193^{+0.0093}_{-0.0057}$
	Exp. cross-feed	$0.0040^{+0.0011}_{-0.0012}$	$0.00195^{+0.00035}_{-0.00032}$	$0.00155^{+0.00025}_{-0.00022}$	$0.00138^{+0.00019}_{-0.00016}$	$0.00159^{+0.00021}_{-0.00019}$	$0.00165^{+0.00021}_{-0.00018}$	$0.00188^{+0.00024}_{-0.00021}$	$0.00196^{+0.00038}_{-0.00034}$
	Exp. signal	$0.0119^{+0.0034}_{-0.0035}$	$0.00573^{+0.00109}_{-0.00096}$	$0.00454^{+0.00078}_{-0.00070}$	$0.00405^{+0.00059}_{-0.00053}$	$0.00470^{+0.00066}_{-0.00061}$	$0.00487^{+0.00065}_{-0.00059}$	$0.00555^{+0.00077}_{-0.00069}$	$0.0057^{+0.0012}_{-0.0010}$
	Observed	110	1	1	1	1	0	0	0
5254 – 5266	Exp. comb. bkg	$203.5^{+4.4}_{-4.4}$	$6.3^{+1.2}_{-1.2}$	$1.69^{+0.45}_{-0.43}$	$0.48^{+0.13}_{-0.12}$	$0.36^{+0.12}_{-0.10}$	$0.115^{+0.073}_{-0.055}$	$0.172^{+0.079}_{-0.062}$	$0.058^{+0.056}_{-0.030}$
	Exp. peak. bkg	$0.104^{+0.062}_{-0.040}$	$0.053^{+0.025}_{-0.016}$	$0.042^{+0.019}_{-0.012}$	$0.038^{+0.016}_{-0.011}$	$0.043^{+0.018}_{-0.013}$	$0.041^{+0.017}_{-0.012}$	$0.042^{+0.018}_{-0.012}$	$0.040^{+0.019}_{-0.012}$
	Exp. cross-feed	$0.0091^{+0.0026}_{-0.0026}$	$0.00439^{+0.00079}_{-0.00071}$	$0.00348^{+0.00057}_{-0.00050}$	$0.00310^{+0.00042}_{-0.00037}$	$0.00359^{+0.00047}_{-0.00042}$	$0.00373^{+0.00046}_{-0.00041}$	$0.00425^{+0.00055}_{-0.00048}$	$0.00441^{+0.00085}_{-0.00077}$
	Exp. signal	$0.037^{+0.011}_{-0.011}$	$0.0177^{+0.0033}_{-0.0029}$	$0.0140^{+0.0024}_{-0.0021}$	$0.0125^{+0.0018}_{-0.0016}$	$0.0145^{+0.0020}_{-0.0019}$	$0.0151^{+0.0020}_{-0.0018}$	$0.0172^{+0.0023}_{-0.0021}$	$0.0178^{+0.0035}_{-0.0032}$
	Observed	203	11	3	0	0	0	0	0
5266 – 5302	Exp. comb. bkg	601^{+13}_{-13}	$18.4^{+3.6}_{-3.5}$	$5.0^{+1.3}_{-1.3}$	$1.40^{+0.38}_{-0.35}$	$1.06^{+0.34}_{-0.30}$	$0.33^{+0.21}_{-0.15}$	$0.48^{+0.22}_{-0.18}$	$0.163^{+0.158}_{-0.086}$
	Exp. peak. bkg	$0.256^{+0.143}_{-0.096}$	$0.130^{+0.059}_{-0.036}$	$0.102^{+0.045}_{-0.027}$	$0.092^{+0.039}_{-0.024}$	$0.104^{+0.043}_{-0.027}$	$0.099^{+0.040}_{-0.026}$	$0.102^{+0.042}_{-0.027}$	$0.097^{+0.044}_{-0.027}$
	Exp. cross-feed	$0.043^{+0.012}_{-0.012}$	$0.0207^{+0.0037}_{-0.0033}$	$0.0164^{+0.0026}_{-0.0024}$	$0.0146^{+0.0020}_{-0.0017}$	$0.0169^{+0.0022}_{-0.0020}$	$0.0175^{+0.0021}_{-0.0019}$	$0.0200^{+0.0026}_{-0.0022}$	$0.0208^{+0.0040}_{-0.0036}$
	Exp. signal	$0.165^{+0.048}_{-0.049}$	$0.080^{+0.015}_{-0.013}$	$0.0631^{+0.0106}_{-0.0096}$	$0.0563^{+0.0080}_{-0.0072}$	$0.0652^{+0.0091}_{-0.0083}$	$0.0676^{+0.0089}_{-0.0081}$	$0.0771^{+0.0105}_{-0.0095}$	$0.080^{+0.016}_{-0.014}$
	Observed	578	18	5	2	1	1	1	0
5302 – 5314	Exp. comb. bkg	$197.1^{+4.1}_{-4.1}$	$6.0^{+1.2}_{-1.1}$	$1.62^{+0.42}_{-0.41}$	$0.46^{+0.12}_{-0.11}$	$0.348^{+0.108}_{-0.096}$	$0.111^{+0.065}_{-0.050}$	$0.152^{+0.067}_{-0.057}$	$0.051^{+0.049}_{-0.027}$
	Exp. peak. bkg	$0.048^{+0.021}_{-0.017}$	$0.0242^{+0.0073}_{-0.0049}$	$0.0190^{+0.0055}_{-0.0035}$	$0.0169^{+0.0049}_{-0.0026}$	$0.0190^{+0.0055}_{-0.0029}$	$0.0179^{+0.0052}_{-0.0026}$	$0.0186^{+0.0054}_{-0.0028}$	$0.0181^{+0.0056}_{-0.0038}$
	Exp. cross-feed	$0.0277^{+0.0078}_{-0.0080}$	$0.0134^{+0.0024}_{-0.0021}$	$0.0106^{+0.0017}_{-0.0015}$	$0.0094^{+0.0013}_{-0.0011}$	$0.0109^{+0.0014}_{-0.0013}$	$0.0113^{+0.0014}_{-0.0012}$	$0.0129^{+0.0017}_{-0.0014}$	$0.0134^{+0.0026}_{-0.0023}$
	Exp. signal	$0.037^{+0.011}_{-0.011}$	$0.0177^{+0.0033}_{-0.0029}$	$0.0140^{+0.0024}_{-0.0021}$	$0.0125^{+0.0018}_{-0.0016}$	$0.0145^{+0.0020}_{-0.0018}$	$0.0151^{+0.0020}_{-0.0018}$	$0.0172^{+0.0024}_{-0.0021}$	$0.0178^{+0.0035}_{-0.0032}$
	Observed	177	3	1	0	0	0	0	0
5314 – 5320	Exp. comb. bkg	$97.9^{+2.0}_{-2.0}$	$2.99^{+0.57}_{-0.57}$	$0.81^{+0.21}_{-0.20}$	$0.227^{+0.058}_{-0.055}$	$0.173^{+0.053}_{-0.047}$	$0.055^{+0.031}_{-0.025}$	$0.074^{+0.033}_{-0.028}$	$0.025^{+0.024}_{-0.013}$
	Exp. peak. bkg	$0.0175^{+0.0068}_{-0.0059}$	$0.0088^{+0.0021}_{-0.0016}$	$0.0069^{+0.0015}_{-0.0012}$	$0.00620^{+0.00118}_{-0.00094}$	$0.0070^{+0.0013}_{-0.0010}$	$0.00657^{+0.00115}_{-0.00095}$	$0.00683^{+0.00126}_{-0.00100}$	$0.0066^{+0.0016}_{-0.0013}$
	Exp. cross-feed	$0.0219^{+0.0063}_{-0.0064}$	$0.0106^{+0.0020}_{-0.0017}$	$0.0084^{+0.0014}_{-0.0012}$	$0.00745^{+0.00108}_{-0.00093}$	$0.0086^{+0.0012}_{-0.0011}$	$0.0090^{+0.0012}_{-0.0011}$	$0.0102^{+0.0014}_{-0.0012}$	$0.0106^{+0.0021}_{-0.0019}$
	Exp. signal	$0.0119^{+0.0034}_{-0.0035}$	$0.00573^{+0.00108}_{-0.00096}$	$0.00454^{+0.00078}_{-0.00070}$	$0.00405^{+0.00059}_{-0.00052}$	$0.00469^{+0.00066}_{-0.00061}$	$0.00487^{+0.00066}_{-0.00060}$	$0.00555^{+0.00078}_{-0.00069}$	$0.0057^{+0.0012}_{-0.0010}$
	Observed	92	0	1	0	0	0	0	0
5320 – 5332	Exp. comb. bkg	$194.7^{+4.1}_{-4.0}$	$5.9^{+1.1}_{-1.1}$	$1.60^{+0.41}_{-0.40}$	$0.45^{+0.11}_{-0.11}$	$0.343^{+0.102}_{-0.092}$	$0.110^{+0.061}_{-0.048}$	$0.144^{+0.063}_{-0.055}$	$0.049^{+0.047}_{-0.026}$
	Exp. peak. bkg	$0.0257^{+0.0104}_{-0.0087}$	$0.0130^{+0.0032}_{-0.0028}$	$0.0103^{+0.0024}_{-0.0021}$	$0.0093^{+0.0019}_{-0.0019}$	$0.0104^{+0.0021}_{-0.0021}$	$0.0099^{+0.0019}_{-0.0020}$	$0.0102^{+0.0020}_{-0.0021}$	$0.0097^{+0.0025}_{-0.0021}$
	Exp. cross-feed	$0.089^{+0.025}_{-0.026}$	$0.0427^{+0.0079}_{-0.0070}$	$0.0339^{+0.0056}_{-0.0050}$	$0.0302^{+0.0042}_{-0.0037}$	$0.0350^{+0.0048}_{-0.0042}$	$0.0363^{+0.0046}_{-0.0041}$	$0.0414^{+0.0055}_{-0.0049}$	$0.0429^{+0.0084}_{-0.0075}$
	Exp. signal	$0.0136^{+0.0040}_{-0.0040}$	$0.0065^{+0.0013}_{-0.0011}$	$0.00519^{+0.00093}_{-0.00081}$	$0.00463^{+0.00070}_{-0.00062}$	$0.00537^{+0.00080}_{-0.00072}$	$0.00557^{+0.00079}_{-0.00071}$	$0.00635^{+0.00093}_{-0.00083}$	$0.0066^{+0.0014}_{-0.0012}$
	Observed	182	3	2	0	0	0	1	0
5332 – 5344	Exp. comb. bkg	$193.2^{+4.0}_{-4.0}$	$5.9^{+1.1}_{-1.1}$	$1.58^{+0.40}_{-0.39}$	$0.45^{+0.11}_{-0.11}$	$0.339^{+0.101}_{-0.089}$	$0.109^{+0.058}_{-0.048}$	$0.140^{+0.060}_{-0.054}$	$0.048^{+0.045}_{-0.026}$
	Exp. peak. bkg	$0.0172^{+0.0092}_{-0.0064}$	$0.0089^{+0.0034}_{-0.0027}$	$0.0071^{+0.0026}_{-0.0021}$	$0.0064^{+0.0022}_{-0.0019}$	$0.0071^{+0.0025}_{-0.0022}$	$0.0067^{+0.0023}_{-0.0021}$	$0.0070^{+0.0024}_{-0.0021}$	$0.0067^{+0.0026}_{-0.0020}$
	Exp. cross-feed	$0.192^{+0.052}_{-0.055}$	$0.093^{+0.016}_{-0.014}$	$0.073^{+0.011}_{-0.010}$	$0.0653^{+0.0079}_{-0.0071}$	$0.0756^{+0.0090}_{-0.0080}$	$0.0785^{+0.0086}_{-0.0077}$	$0.0895^{+0.0102}_{-0.0092}$	$0.093^{+0.017}_{-0.016}$
	Exp. signal	$0.0052^{+0.0016}_{-0.0016}$	$0.00249^{+0.00053}_{-0.00046}$	$0.00198^{+0.00039}_{-0.00034}$	$0.00176^{+0.00030}_{-0.00027}$	$0.00204^{+0.00035}_{-0.00030}$	$0.00212^{+0.00035}_{-0.00030}$	$0.00242^{+0.00041}_{-0.00035}$	$0.00250^{+0.00056}_{-0.00049}$
	Observed	186	3	1	1	0	0	0	0

Table 5.25: $B^0 \to \mu^+\mu^-$: Expected combinatorial background events, expected peaking ($B^0_{(s)} \to h^+h'^-$) background events, expected cross-feed events from $B^0_s \to \mu^+\mu^-$ assuming the SM branching fraction and expected $B^0 \to \mu^+\mu^-$ signal events assuming the SM branching fraction.

Invariant mass [MeV/c^2]		BDT							
		0.0 – 0.25	0.25 – 0.4	0.4 – 0.5	0.5 – 0.6	0.6 – 0.7	0.7 – 0.8	0.8 – 0.9	0.9 – 1.0
5312 – 5324	Exp. comb. bkg	$195.7^{+4.1}_{-4.1}$	$6.0^{+1.2}_{-1.1}$	$1.61^{+0.41}_{-0.40}$	$0.45^{+0.12}_{-0.11}$	$0.345^{+0.105}_{-0.094}$	$0.110^{+0.063}_{-0.049}$	$0.147^{+0.065}_{-0.056}$	$0.050^{+0.048}_{-0.027}$
	Exp. peak. bkg	$0.035^{+0.016}_{-0.012}$	$0.0177^{+0.0058}_{-0.0043}$	$0.0139^{+0.0044}_{-0.0033}$	$0.0125^{+0.0039}_{-0.0028}$	$0.0140^{+0.0043}_{-0.0032}$	$0.0132^{+0.0041}_{-0.0030}$	$0.0137^{+0.0042}_{-0.0032}$	$0.0132^{+0.0044}_{-0.0033}$
	Exp. signal	$0.050^{+0.014}_{-0.015}$	$0.0240^{+0.0046}_{-0.0040}$	$0.0191^{+0.0033}_{-0.0029}$	$0.0170^{+0.0025}_{-0.0021}$	$0.0197^{+0.0028}_{-0.0024}$	$0.0204^{+0.0027}_{-0.0024}$	$0.0233^{+0.0033}_{-0.0028}$	$0.0241^{+0.0049}_{-0.0043}$
	Observed	188	2	3	0	0	0	1	0
5324 – 5336	Exp. comb. bkg	$194.2^{+4.1}_{-4.0}$	$5.9^{+1.1}_{-1.1}$	$1.59^{+0.41}_{-0.40}$	$0.45^{+0.11}_{-0.11}$	$0.341^{+0.102}_{-0.091}$	$0.109^{+0.060}_{-0.048}$	$0.142^{+0.062}_{-0.055}$	$0.049^{+0.046}_{-0.026}$
	Exp. peak. bkg	$0.0237^{+0.0093}_{-0.0083}$	$0.0120^{+0.0028}_{-0.0024}$	$0.0094^{+0.0020}_{-0.0018}$	$0.0085^{+0.0016}_{-0.0014}$	$0.0095^{+0.0018}_{-0.0016}$	$0.0090^{+0.0016}_{-0.0015}$	$0.0094^{+0.0018}_{-0.0016}$	$0.0090^{+0.0022}_{-0.0019}$
	Exp. signal	$0.120^{+0.033}_{-0.035}$	$0.0579^{+0.0103}_{-0.0092}$	$0.0459^{+0.0073}_{-0.0065}$	$0.0409^{+0.0053}_{-0.0047}$	$0.0474^{+0.0060}_{-0.0053}$	$0.0492^{+0.0058}_{-0.0052}$	$0.0560^{+0.0070}_{-0.0062}$	$0.0582^{+0.0111}_{-0.0100}$
	Observed	185	4	1	0	0	0	0	0
5336 – 5342	Exp. comb. bkg	$96.5^{+2.0}_{-2.0}$	$2.94^{+0.56}_{-0.56}$	$0.79^{+0.20}_{-0.20}$	$0.223^{+0.055}_{-0.053}$	$0.169^{+0.051}_{-0.044}$	$0.054^{+0.029}_{-0.024}$	$0.069^{+0.030}_{-0.027}$	$0.024^{+0.022}_{-0.013}$
	Exp. peak. bkg	$0.0092^{+0.0038}_{-0.0033}$	$0.0046^{+0.0012}_{-0.0010}$	$0.00365^{+0.00090}_{-0.00077}$	$0.00328^{+0.00075}_{-0.00063}$	$0.00368^{+0.00083}_{-0.00070}$	$0.00347^{+0.00077}_{-0.00064}$	$0.00361^{+0.00081}_{-0.00068}$	$0.00347^{+0.00095}_{-0.00082}$
	Exp. signal	$0.103^{+0.028}_{-0.030}$	$0.0494^{+0.0083}_{-0.0076}$	$0.0392^{+0.0059}_{-0.0053}$	$0.0349^{+0.0042}_{-0.0038}$	$0.0404^{+0.0047}_{-0.0042}$	$0.0419^{+0.0045}_{-0.0040}$	$0.0478^{+0.0054}_{-0.0048}$	$0.0496^{+0.0090}_{-0.0083}$
	Observed	82	1	0	1	0	0	0	0
5342 – 5354	Exp. comb. bkg	$191.8^{+4.0}_{-4.0}$	$5.8^{+1.1}_{-1.1}$	$1.56^{+0.40}_{-0.39}$	$0.44^{+0.11}_{-0.10}$	$0.336^{+0.099}_{-0.088}$	$0.108^{+0.056}_{-0.047}$	$0.135^{+0.058}_{-0.053}$	$0.047^{+0.043}_{-0.025}$
	Exp. peak. bkg	$0.0136^{+0.0080}_{-0.0055}$	$0.0070^{+0.0032}_{-0.0022}$	$0.0055^{+0.0024}_{-0.0017}$	$0.0050^{+0.0021}_{-0.0015}$	$0.0056^{+0.0023}_{-0.0017}$	$0.0054^{+0.0021}_{-0.0017}$	$0.0055^{+0.0022}_{-0.0017}$	$0.0052^{+0.0024}_{-0.0017}$
	Exp. signal	$0.312^{+0.084}_{-0.089}$	$0.150^{+0.025}_{-0.023}$	$0.119^{+0.018}_{-0.016}$	$0.106^{+0.012}_{-0.011}$	$0.122^{+0.014}_{-0.012}$	$0.127^{+0.013}_{-0.012}$	$0.145^{+0.016}_{-0.014}$	$0.150^{+0.027}_{-0.025}$
	Observed	167	2	3	1	0	0	0	0
5354 – 5390	Exp. comb. bkg	566^{+12}_{-12}	$17.2^{+3.3}_{-3.3}$	$4.6^{+1.2}_{-1.1}$	$1.30^{+0.31}_{-0.30}$	$0.99^{+0.28}_{-0.26}$	$0.32^{+0.15}_{-0.14}$	$0.38^{+0.16}_{-0.15}$	$0.133^{+0.121}_{-0.073}$
	Exp. peak. bkg	$0.021^{+0.028}_{-0.011}$	$0.0137^{+0.0093}_{-0.0079}$	$0.0113^{+0.0066}_{-0.0067}$	$0.0110^{+0.0049}_{-0.0068}$	$0.0125^{+0.0054}_{-0.0077}$	$0.0121^{+0.0047}_{-0.0076}$	$0.0124^{+0.0052}_{-0.0077}$	$0.0098^{+0.0075}_{-0.0055}$
	Exp. signal	$1.37^{+0.37}_{-0.39}$	$0.66^{+0.11}_{-0.10}$	$0.523^{+0.077}_{-0.070}$	$0.466^{+0.055}_{-0.049}$	$0.539^{+0.061}_{-0.055}$	$0.559^{+0.059}_{-0.052}$	$0.638^{+0.071}_{-0.062}$	$0.66^{+0.12}_{-0.11}$
	Observed	557	22	3	2	0	2	0	1
5390 – 5402	Exp. comb. bkg	$185.8^{+3.8}_{-3.8}$	$5.6^{+1.1}_{-1.1}$	$1.50^{+0.37}_{-0.37}$	$0.424^{+0.098}_{-0.097}$	$0.323^{+0.090}_{-0.081}$	$0.103^{+0.048}_{-0.043}$	$0.117^{+0.055}_{-0.049}$	$0.042^{+0.037}_{-0.024}$
	Exp. peak. bkg	$0.0034^{+0.0091}_{-0.0022}$	$0.0035^{+0.0024}_{-0.0028}$	$0.0029^{+0.0017}_{-0.0023}$	$0.0028^{+0.0013}_{-0.0023}$	$0.0032^{+0.0014}_{-0.0026}$	$0.0031^{+0.0012}_{-0.0026}$	$0.0032^{+0.0013}_{-0.0026}$	$0.0025^{+0.0019}_{-0.0020}$
	Exp. signal	$0.311^{+0.084}_{-0.089}$	$0.150^{+0.025}_{-0.023}$	$0.119^{+0.018}_{-0.016}$	$0.106^{+0.012}_{-0.011}$	$0.122^{+0.014}_{-0.012}$	$0.127^{+0.013}_{-0.012}$	$0.145^{+0.016}_{-0.014}$	$0.150^{+0.027}_{-0.025}$
	Observed	219	4	2	1	0	0	0	0
5402 – 5408	Exp. comb. bkg	$92.3^{+1.9}_{-1.9}$	$2.78^{+0.53}_{-0.53}$	$0.74^{+0.18}_{-0.18}$	$0.210^{+0.048}_{-0.048}$	$0.160^{+0.045}_{-0.039}$	$0.051^{+0.023}_{-0.021}$	$0.057^{+0.028}_{-0.024}$	$0.021^{+0.018}_{-0.012}$
	Exp. peak. bkg	$0.0015^{+0.0041}_{-0.0010}$	$0.0016^{+0.0011}_{-0.0013}$	$0.00129^{+0.00076}_{-0.00108}$	$0.00126^{+0.00056}_{-0.00107}$	$0.00142^{+0.00062}_{-0.00121}$	$0.00138^{+0.00054}_{-0.00118}$	$0.00141^{+0.00059}_{-0.00120}$	$0.00112^{+0.00085}_{-0.00092}$
	Exp. signal	$0.103^{+0.028}_{-0.030}$	$0.0494^{+0.0084}_{-0.0075}$	$0.0391^{+0.0059}_{-0.0053}$	$0.0349^{+0.0042}_{-0.0037}$	$0.0404^{+0.0047}_{-0.0042}$	$0.0419^{+0.0045}_{-0.0040}$	$0.0478^{+0.0054}_{-0.0048}$	$0.0496^{+0.0091}_{-0.0082}$
	Observed	74	2	0	0	0	0	0	0
5408 – 5420	Exp. comb. bkg	$183.6^{+3.7}_{-3.7}$	$5.5^{+1.0}_{-1.0}$	$1.48^{+0.37}_{-0.36}$	$0.418^{+0.095}_{-0.094}$	$0.318^{+0.089}_{-0.078}$	$0.101^{+0.046}_{-0.042}$	$0.110^{+0.056}_{-0.046}$	$0.040^{+0.035}_{-0.023}$
	Exp. peak. bkg	$0.0029^{+0.0079}_{-0.0021}$	$0.0030^{+0.0020}_{-0.0026}$	$0.0025^{+0.0015}_{-0.0021}$	$0.0024^{+0.0011}_{-0.0021}$	$0.0027^{+0.0012}_{-0.0024}$	$0.0027^{+0.0010}_{-0.0023}$	$0.0027^{+0.0011}_{-0.0024}$	$0.0022^{+0.0016}_{-0.0018}$
	Exp. signal	$0.121^{+0.033}_{-0.035}$	$0.0580^{+0.0103}_{-0.0092}$	$0.0459^{+0.0073}_{-0.0065}$	$0.0409^{+0.0053}_{-0.0047}$	$0.0474^{+0.0060}_{-0.0054}$	$0.0491^{+0.0058}_{-0.0052}$	$0.0560^{+0.0069}_{-0.0062}$	$0.0581^{+0.0111}_{-0.0099}$
	Observed	176	0	0	0	0	0	0	0
5420 – 5432	Exp. comb. bkg	$182.1^{+3.7}_{-3.7}$	$5.5^{+1.0}_{-1.0}$	$1.46^{+0.36}_{-0.36}$	$0.413^{+0.093}_{-0.093}$	$0.317^{+0.087}_{-0.077}$	$0.100^{+0.044}_{-0.042}$	$0.106^{+0.056}_{-0.044}$	$0.039^{+0.033}_{-0.022}$
	Exp. peak. bkg	$0.0024^{+0.0067}_{-0.0018}$	$0.0026^{+0.0017}_{-0.0022}$	$0.0021^{+0.0012}_{-0.0018}$	$0.00207^{+0.00092}_{-0.00182}$	$0.0023^{+0.0010}_{-0.0021}$	$0.00227^{+0.00088}_{-0.00200}$	$0.00232^{+0.00097}_{-0.00204}$	$0.0018^{+0.0014}_{-0.0016}$
	Exp. signal	$0.048^{+0.014}_{-0.014}$	$0.0231^{+0.0047}_{-0.0040}$	$0.0183^{+0.0034}_{-0.0030}$	$0.0163^{+0.0026}_{-0.0023}$	$0.0189^{+0.0030}_{-0.0026}$	$0.0196^{+0.0029}_{-0.0026}$	$0.0224^{+0.0034}_{-0.0030}$	$0.0231^{+0.0050}_{-0.0043}$
	Observed	170	2	0	1	1	0	0	0

Table 5.26: $B^0_s \to \mu^+\mu^-$: Expected combinatorial background events, expected peaking ($B^0_{(s)} \to h^+h'^-$) background events and expected signal events assuming the SM branching fraction prediction.

$B^0 \to \mu^+\mu^-$	at 90% CL	at 95% CL	CL_b
expected limits: bkg-only	0.91×10^{-9}	1.1×10^{-9}	–
observed limits	$\mathbf{0.81 \times 10^{-9}}$	$\mathbf{1.0 \times 10^{-9}}$	0.40

Table 5.27: Expected and observed limits for the $B^0 \to \mu^+\mu^-$ channel.

$B^0_s \to \mu^+\mu^-$	at 90% CL	at 95% CL	CL_b
expected limits: bkg-only	2.8×10^{-9}	3.4×10^{-9}	–
expected limits: bkg+SM	6.3×10^{-9}	7.3×10^{-9}	–
observed limits	$\mathbf{3.8 \times 10^{-9}}$	$\mathbf{4.5 \times 10^{-9}}$	0.82

Table 5.28: Expected and observed limits for the $B^0_s \to \mu^+\mu^-$ channel

5.11 BR results

Several different methods are possible for measuring a $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$. These include a comparison between data and signal and background expectations in bins of mass and BDT, a likelihood fit to the unbinned mass projections simultaneously in the 8 different BDT bins, and a two-dimensional unbinned fit in the mass-BDT plane. These methods are discussed in detail in [173].

The simultaneous likelihood fit in the 8 different BDT bins will be briefly explained here, also showing the results obtained in the analysed 2011 dataset. It has to be clarified that the obtention of the $\mathcal{B}$ is done even if no significant signal excess has been observed.

This fit is performed simultaneously in the 8 BDT bins, with the signal constrained through the BDT fractions calibrated on the $B^0_{(s)} \to h^+h'^-$ sample (section 5.8). The same occurs for the misidentified $B^0_{(s)} \to h^+h'^-$ events, with BDT and mass PDFs as described in section 5.6. The uncertainties on the parameters defining these PDFs are included in the fit as Gaussian constraints. The combinatorial background is instead evaluated for each BDT bin by including the mass sidebands in the fit range. The mass PDF assumed for this component is an exponential. The $\mathcal{B}$ is extracted directly from the fit, by defining $\mathrm{BR}/\alpha_{B^0_s\to\mu^+\mu^-}$ as the normalisation for the signal PDF.

The mass fits for the 8 BDT bins are shown in figure 5.55 for the analysed dataset. Moreover, and in order to obtain the errors in the $\mathcal{B}$ estimation, the resulting profile likelihood is also shown in the B^0_s case in figure 5.56. The fitted $\mathcal{B}$s are $\mathcal{B}(\mathrm{B^0_s} \to \mu^+\mu^-) = (0.79^{+1.68}_{-1.17}) \times 10^{-9}$, and $\mathcal{B}(\mathrm{B^0} \to \mu^+\mu^-) = (0.06^{+5.46}_{-4.27}) \times 10^{-10}$. The quoted errors include both the statistical fluctuations and the uncertainties from the knowledge of the systematics, which as said are included in the fit through Gaussian constraints.

Furthermore, by taking the difference in the fit results using a double exponential for the combinatorial background mass PDF instead of a single exponential, an additional systematic error of $\pm 1.95 \times 10^{-10}$ and $\pm 0.46 \times 10^{-9}$ has to be added in quadrature to the B^0 and B^0_s results, respectively.

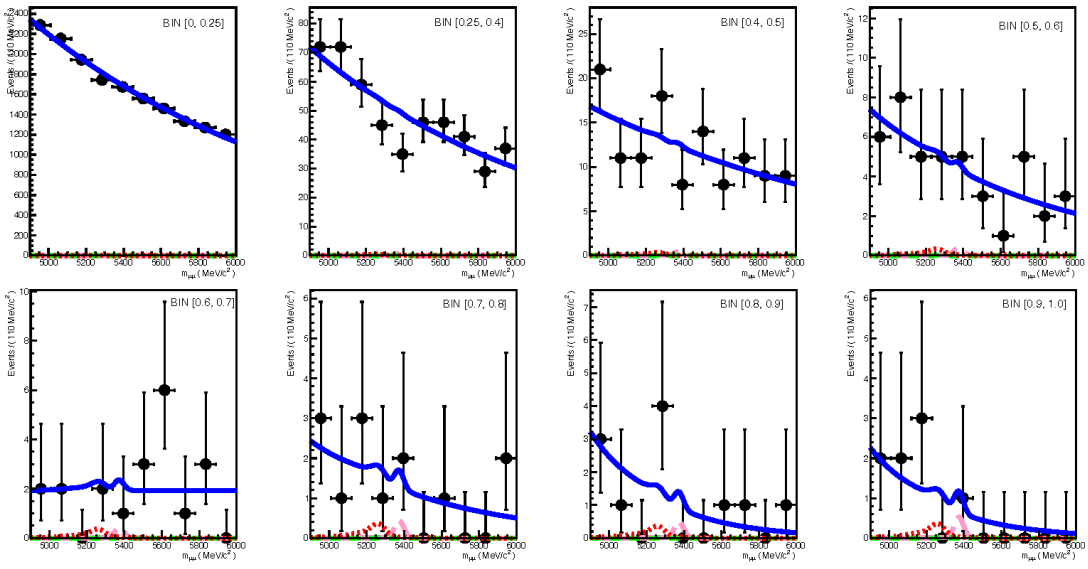

Figure 5.55: Likelihood mass fit in the 8 BDT bins, for the unblinded sample: blue line is the total fit result, pink is $B^0_s \to \mu^+\mu^-$ signal, red is $B^0_{(s)} \to h^+h'^-$ misID.

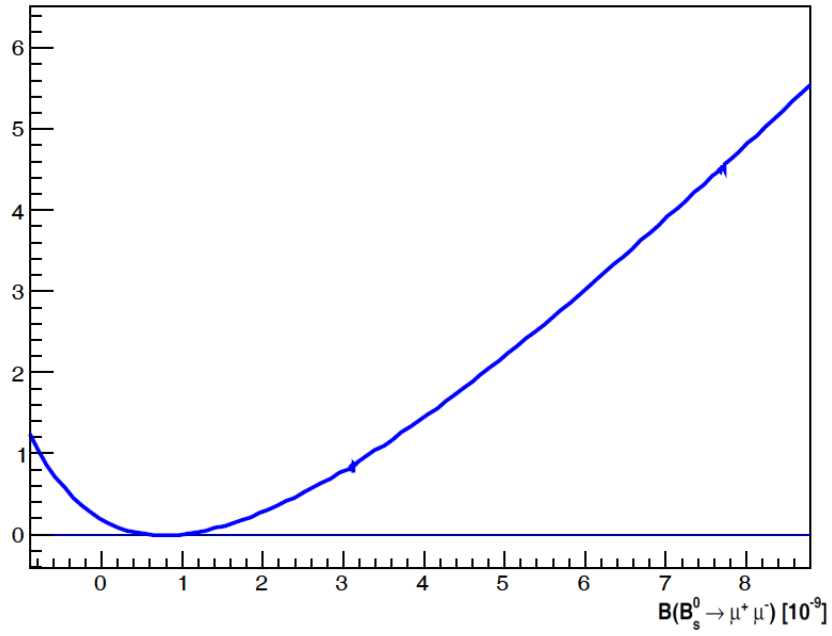

Figure 5.56: Unblinded data sample: likelihood profile obtained with the simultaneous fit.

5.A Theory implications of the result

The results shown for the upper limits on $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and specially $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ set important bounds for the NP searches, according to what it was explained in section 2.2.3. In this way, out of the scenarios included in table 2.2, the first ($\mathcal{B}(B^0_s \to \mu^+\mu^-) \gg SM$) can be already discarded. Similarly, the NP phase space region where $\tan\beta$ is large is clearly disfavoured by the $B^0_s \to \mu^+\mu^-$ result.

Given the interest of this result, the theory community has reacted promptly to study the implications of these limits in several NP models. A few examples are shown next:

- A very wide review of these implications can be found in [45]. Figure 5.57 shows how both results in $B^0_{(s)} \to \mu^+\mu^-$ severely constraint the phase space of various NP models, including MFV, SM4 and different examples of SF.

- The impact of the limit in $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ in the phase space of NUHM, CMSSM and CNMSSM, all of them explained in section 2.2.2, can be found in [136]. This can be seen in figure 5.58 (5.59) for NUHM and CNMSSM (CMSSM).

- Reference [185] gathers the implications on CMSSM and NUHM1 of the results of several LHC measurements using the data taken during 2010, 2011 and 2012 together with the constraint on the spin-independent dark matter scattering cross section imposed by 225 live days of XENON100 data [186]. Among the former is the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$. Figure 5.60 shows the particular impact of the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ limit in the phase space of both models. Similar studies have been performed in [187] for CMSSM and NUHM1 and in [188] for CMSSM.

- Interesting SUSY bounds with previous LHCb limits on $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ can be also found in [189].

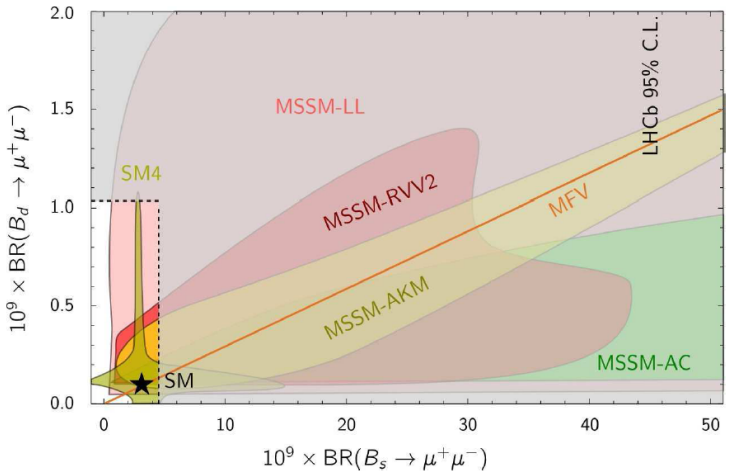

Figure 5.57: Bounds to the phase space in different NP models by the limits in $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$, extracted from [45, 180]. Models in the plot include MFV, SM4 and several examples of SF models (the acronyms in the plot stand for the models by Agashe and Carone, AC, [181]; Ross, Velasco-Sevilla and Vives, RVV2, [182]; Antusch, King and Malinsky, AKM, [183]; and a model with left-handed currents only, LL, [184]). The SM is marked with a star.

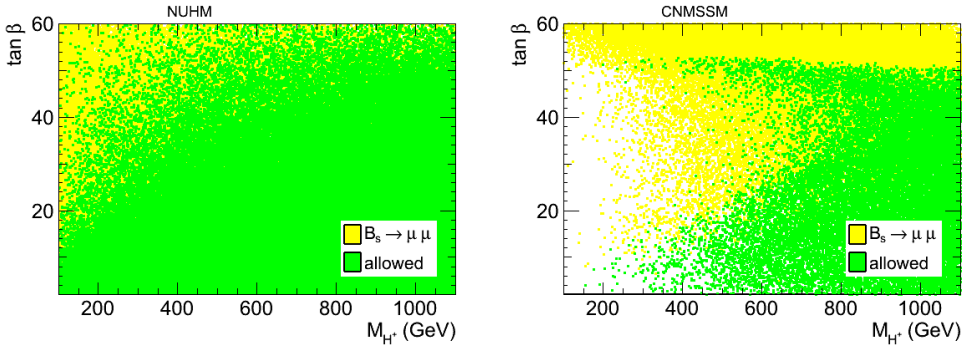

Figure 5.58: Bounds to the phase space of NUHM (left) and CNMSSM (right) from the limit in $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$, extracted from [136]. The x axis represents the mass of the charged Higgs in the model (H^+), while the y axis the $\tan\beta$ parameter explained in section 2.2.2. Green points correspond to the allowed region (in the foreground), while yellow to the one excluded by the limit in 5.28 (in the background).

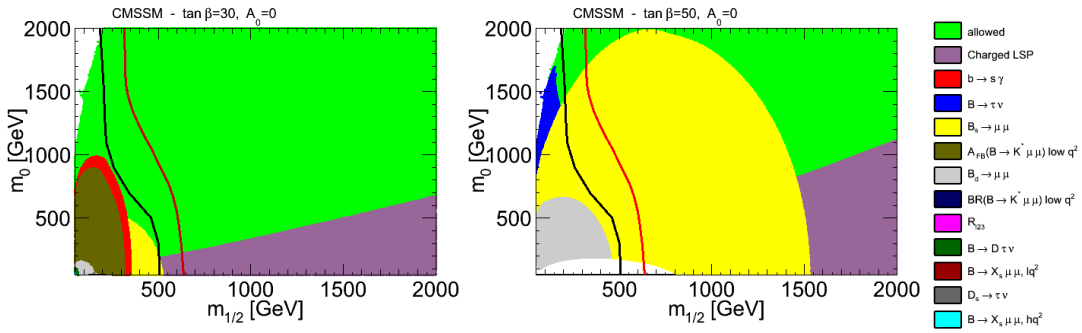

Figure 5.59: Bounds to the phase space of CMSSM from the limit in $\mathcal{B}(B^0 \to \mu^+\mu^-)$, $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ and other measurements, extracted from [136]. The plots shown correspond to $A_0 = 0$ and $\tan\beta = 30$ (left), $\tan\beta = 50$ (right). The x axis represents the $m_{1/2}$ parameter (universal gaugino mass), the y axis the m_0 parameter (universal mass of the scalars), while A_0 is the universal trilinear coupling and $\tan\beta$ was explained in section 2.2.2. The colour legend can be found in the right hand side of figure (grey is the area excluded by $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and yellow by $\mathcal{B}(B^0_s \to \mu^+\mu^-)$). The black line corresponds to the CMS exclusion limit with 1.1 fb^{-1} of data [190] and the red line to the CMS exclusion limit with 4.4 fb^{-1} of data [191].

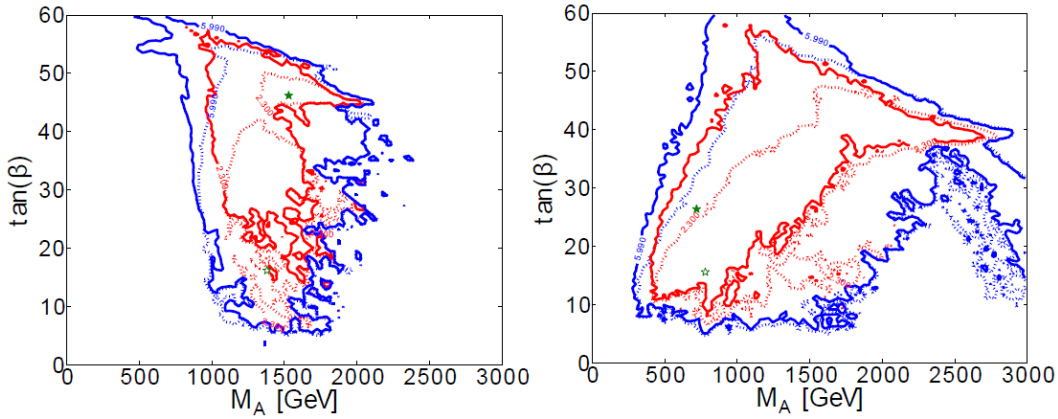

Figure 5.60: Bounds to the phase space of CMSSM (left) and NUHM1 (right) extracted from [185]. The (M_A, $\tan\beta$) planes are displayed, showing the effects of the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ limit. M_A is the mass of the MSSM Higgs pseudoscalar boson, while $\tan\beta$ was explained in section 2.2.2. The bounds including all the available data are represented by an open green star and dashed blue and red lines for the 68 and 95% CL contours, while the results after dropping the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ constraint are represented by closed stars and solid contours.

5.B Combination with other experiments

Given the importance of the $B^0_{(s)} \to \mu^+\mu^-$ decays, these have also been searched for by other two LHC experiments, ATLAS and CMS (see section 3.1), using the 2011 pp collisions taken at $\sqrt{s}$=7 TeV . In particular, ATLAS has searched for the $B^0_s \to \mu^+\mu^-$ decay with 2.4 fb^{-1} [192] and CMS has searched for both $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ with 5 fb^{-1} [193]. Neither ATLAS nor CMS have found any significant signal excess so, similarly to LHCb, they have set upper limits in the $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$. Table 5.29 shows the result obtained by CMS for the upper limit on $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and compares it with the one obtained by LHCb. In the same way, table 5.30 gathers the results obtained by ATLAS and CMS for the upper limit on $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ and compares them with the result obtained by LHCb. LHCb result is world best in both cases.

Experiment	Dataset	Limit
CMS	5 fb^{-1}	18×10^{-10}
LHCb	1 fb^{-1}	10×10^{-10}

Table 5.29: Upper limits on the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ at 95% CL obtained by different experiments using pp collisions at $\sqrt{s}$=7 TeV during year 2011.

Experiment	Dataset	Limit
ATLAS	2.4 fb^{-1}	22×10^{-9}
CMS	5 fb^{-1}	7.7×10^{-9}
LHCb	1 fb^{-1}	4.5×10^{-9}

Table 5.30: Upper limits on the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ at 95% CL obtained by different experiments using pp collisions at $\sqrt{s}$=7 TeV during year 2011.

The CL$_s$ method used to obtain the LHCb result has also been used in [194] to combine the results from the three mentioned experiments. In the case of LHCb, the result obtained with 2010 data [148] is also included.

Figure 5.61 shows the expected and observed CL$_s$ value derived from CMS and the two LHCb analyses as a function of the assumed $\mathcal{B}(B^0 \to \mu^+\mu^-)$. The expected and measured limits can be found in table 5.31.

As for $B^0_s \to \mu^+\mu^-$, the expected and observed CL$_s$ values derived from ATLAS, CMS, and the two LHCb analyses can be found in 5.62 as a function of the assumed $\mathcal{B}(B^0_s \to \mu^+\mu^-)$. The expected and measured limits can be also found in table 5.31.

Channel	Limit exp. (bkg. only)	Limit exp. (bkg.+SM)	**Limit observed**
$B^0 \to \mu^+\mu^-$ (10^{-10})	7.3	–	**8.1**
$B^0_s \to \mu^+\mu^-$ (10^{-9})	2.3	6.1	**4.2**

Table 5.31: Expected and observed upper limits on the $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$ at 95% CL combining the results of the three experiments (see text for details). In the case of $B^0_s \to \mu^+\mu^-$, the expected limit is shown in the background only and background+SM signal hypotheses.

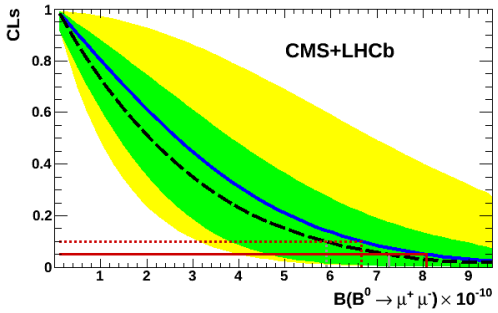

Figure 5.61: Combined CL_s as a function of the assumed branching fraction for $B^0 \to \mu^+\mu^-$. The dashed black curve is the median of the expected CL_s distribution, in absence of signal. The green (yellow) areas cover, for each branching fraction, ± 1 (2) σ intervals around the median. The solid blue curve is the observed CL_s. The upper limits at 90% (95 %) C.L. are indicated by the dotted (solid) horizontal lines in red (dark gray) for the observation and in gray for the expectation.

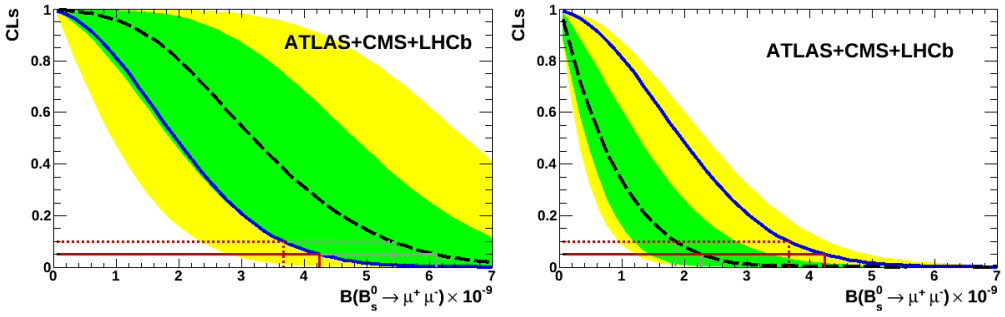

Figure 5.62: Combined CL_s as a function of the assumed branching fraction for $B^0_s \to \mu^+\mu^-$. The dashed black curves are the medians of the expected CL_s distributions, if background and SM signal were observed (left), and in absence of signal (right). The green (yellow) areas cover, for each branching fraction, ± 1 (2) σ intervals around the medians. The solid blue curves are the observed CL_s. The upper limits at 90% (95 %) C.L. are indicated by the dotted (solid) horizontal lines in red (dark gray) for the observation and in gray for the expectation.

6

$K^0_S \to \mu^+\mu^-$

6.1 Introduction

As explained in sections 2.1.4 and 2.2.4, $K^0_S \to \mu^+\mu^-$ is a FCNC decay very suppressed in the SM but with a SD contribution sensitive to NP contributions. The SM prediction on the $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ is $(5.1 \pm 1.5) \times 10^{-12}$ (see equation 2.23).

As for the experimental searches, the best upper limit up to now (almost 40 years old), was found in [195] to be 3.2×10^{-7} at 90 % CL.

This chapter is devoted to the analysis of the search for $K^0_S \to \mu^+\mu^-$, performed, as in the $B^0_{(s)} \to \mu^+\mu^-$ case, in the ~ 1 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV collected by LHCb in 2011.

An overview of the analysis is given next. The $K^0_S \to \mu^+\mu^-$ candidates are selected requiring a very detached Secondary Vertex (SV) with a reconstructed momentum pointing to one of the pp collisions in the event, and an invariant mass in the region of the K^0_S mass.

The selected events are triggered mainly in two different ways: via Single Muon triggers which require a high p_T muon (p_T above 1.3 GeV/c), or via some other particle in the underlying event different from the K^0_S daughters. The events triggered by one of the K^0_S muons are TOS, while those triggered by the underlying event are TIS[1]. The analysis is performed separately in TIS and in TOS events, due to major differences in the signal and background properties of those two categories, which will be explained in the next sections. Most of sensitivity in the search comes from TOS events, but, as it will be seen later, a potential gain of $\sim 10\%$ in the upper limit can be achieved if the TIS events are also included. In addition, the analysis of TIS events is much simpler than that of of TOS events, due to the easier normalisation procedure, and alone can still provide an upper limit one order of magnitude better than current world best in [195].

After the trigger and selection, further background discrimination is achieved by the use of a Boosted Decision Tree (BDT). This BDT is trained using data, which is splitted in two samples so that the BDT applied in each half has been trained in the complementary half. This procedure avoids having any overlap between the training sample and the analysed sample.

The signal candidates are classified in bins of the BDT, and a limit is computed using CL_s method, explained in chapter 5, based on the background and signal expectations and the observed events. The background expectation is obtained from the mass sidebands, and is

[1] There are also events in which the trigger decision relies partially on the signal and partially on the underlying event (TOB events), as well as a small fraction of TOS events from lines different than the single muon. But the contribution of those two cases is small, and very difficult to treat.

composed mainly by combinatorial background and $K_S^0 \to \pi^+\pi^-$ with both pions misidentified as muons. The $K_S^0 \to \pi^+\pi^-$ misid is well separated from the $K_S^0 \to \mu^+\mu^-$ because of the effect of the different mass between muon and pion, but produces a tail that enters in the signal region. Backgrounds from $K_L^0 \to \mu^+\mu^-$, $K_S^0 \to \mu^+\mu^-\gamma$ and $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu$ are found to be negligible for the current luminosity. The signal expectations are obtained using $K_S^0 \to \pi^+\pi^-$. A dedicated selection, prescaled to fit into the LHCb standard rate requirements, has been used, based on the same cuts as for $K_S^0 \to \mu^+\mu^-$ with the exception of muon identification cuts. For the TIS analysis, $K_S^0 \to \pi^+\pi^-$ TIS events are used, whereas for the TOS analysis the simultaneous use of $K_S^0 \to \pi^+\pi^-$ from physics lines and from minimum bias lines is required. Muons from $J/\psi \to \mu^+\mu^-$ are used to get the efficiency of the muon identification and of the muon triggers. MC simulation, properly weighted to match the η, p_T distributions of the K_S^0 in data, is used to calculate efficiency ratios.

6.2 Data and Monte Carlo Samples

6.2.1 Data Sample

As in the case of $B_{(s)}^0 \to \mu^+\mu^-$, the search for $K_S^0 \to \mu^+\mu^-$ has been done using the data collected by LHCb between March and October 2011 at $\sqrt{s} = 7$ Tev, corresponding to ~ 1.02 fb^{-1} of integrated luminosity. Again, the data belong to the Reco12-Stripping17 campaign, whose details were explained in section 5.2.

6.2.2 Monte Carlo samples

The Monte Carlo samples used in this analysis belong mainly to the MC11a, sim05 and sim05a generation. Two samples coming from MC10, sim01 were used too. The specific simulation decfiles[2] written for this analysis are summarised in table 6.1. Other samples used are listed in table 6.2. A brief reminder of the generation/reconstruction conditions is reported below

For the MC11a sim05 and sim05a, the *pp* interactions were simulated assuming a beam energy of 3.5 TeV, an average number of interactions per crossing $\nu = 2.0$, which corresponds to an average number of visible interactions per crossing $\mu = 1.40$. The samples were generated using the GAUDI v22r4 application GAUSS v41r2, based on LHCb v33r0, Pythia 6_425.2, PHOTOS 215.4, LHAPDF 5.8.6., HepMC 2.06.05, and GEANT4 v94r2p1.p02. Then they were reconstructed using Boole v23r1 and Brunel v41r1p1.

For the MC10 sim01, the details have already been explained in section 5.2.

The K_S^0 in the minimum bias samples are generated in the full solid angle. In the exclusive samples, only K_S^0 in the forward direction are passed through the full detector simulation.

For the MC11a samples, the default trigger emulated in *pass-through mode* corresponds to *Trigger Configuration Key* (TCK) 0x40760037. This TCK includes the cuts used in most of 2011 data taking, and removes the prescales existing in several trigger lines. The stripping (stripping17, see sections 6.3 and 5.4) $K_S^0 \to \mu^+\mu^-$ and $K_S^0 \to \pi^+\pi^-$ selections were applied to strip all the samples. These selections are the same used for data.

For the MC10 samples, the trigger was emulated again in *pass-through mode* with a configuration with equivalent cuts to those more extensively used during the 2011 data taking, TCK 0x006d0032. In addition, the same stripping selections used for data were also applied to strip the MC10 samples.

[2]Decfiles are the standard LHCb files including the MC configuration for different event types.

EvtType Nickname	stat	comments
34112100 Ks_mumu	1M	no generator level cuts
34112102 Ks_mumu=TightCut	2M	K^0_S EndVertex(z) $\in [-1, 1]$ m
		K^0_S EndVertex(ρ) < 30 mm
34112103 Ks_mumu=TightCut,pt1.5GeV		K^0_S EndVertex(z) $\in [-1, 1]$ m
		K^0_S EndVertex(ρ) < 30 mm
		K^0_S $p_T > 1.5$ GeV
34512100 Ks_pimunu	100k	no generator level cuts
34512103 Ks_pimunu,m=TightCut	2M	π forced to $\mu\nu$
		π EndVertex(z) $\in [-1, 11.5]$ m
34102101 Ks_pipi=TightCut		K^0_S EndVertex(z) $\in [-1, 1]$ m
34102102 Ks_pipi=TightCut,rho		K^0_S EndVertex(z) $\in [-1, 1]$ m
		K^0_S EndVertex(ρ) < 30 mm
34102103 Ks_pipi=TightCut,pt1.5GeV		K^0_S EndVertex(z) $\in [-1, 1]$ m
		K^0_S EndVertex(ρ) < 30 mm
		K^0_S $p_T > 1.5$ GeV
34512104 Ks_pipi,mm=TightCut	2M	both π forced to $\mu\nu$
		both π EndVertex(z) $\in [-1, 11.5]$ m

Table 6.1: Summary of the Simulation decfiles written for the $K^0_S \to \mu^+\mu^-$ analysis. They all correspond to MC11a.

Channel	Monte Carlo production	Events processed
MinBias	MC11aSim05	210 k
$B^+ \to J/\psi K^+$	MC11aSim05	517 k
$B^+ \to J/\psi K^+$	MC10sim01	2 M
MinBias	MC10sim01	70.5 M

Table 6.2: Monte Carlo samples used in the $K^0_S \to \mu^+\mu^-$ analysis not contained in table 6.1 (i.e, not generated specifically for this analysis). The Monte Carlo production version and the number of processed events are in the second and third columns, respectively.

6.3 Selection

This section is devoted to the description of the initial selections applied both to $K^0_S \to \mu^+\mu^-$ and $K^0_S \to \pi^+\pi^-$. These include the stripping selections and some extra fiducial cuts applied offline to all the candidates.

The stripping selection is designed to be the same between the signal and the control channel, with the exception of the muon identification requirements and the mass windows. This is common to the $B^0_{(s)} \to \mu^+\mu^-$ analysis seen in previous chapter. These selections use the same cuts as in [196]. A summary of the cuts applied can be found in table 6.3.

The search for $K^0_S \to \mu^+\mu^-$ is performed by using only the "long-long" category, i.e., both daughters of the K^0_S being long tracks (see section 3.2.3.2.1). The reason not to use the downstream-downstream candidates arises from the fact that the excellent K^0_S mass resolution achieved with long tracks (at the level of $\sim 3-4$ MeV/c^2), is heavily downgraded for the downstream case. The mass resolution is crucial for the $K^0_S \to \mu^+\mu^-$ analysis, as it allows the separation between the $K^0_S \to \pi^+\pi^-$ with both pions misidentified as muons from the real signal. This will be seen in detail in section 6.6.2. Another reason to discard the downstream-downstream category is the larger influence (with respect to long-long) of the $K^0_L \to \mu^+\mu^-$, which is an irreducible background for $K^0_S \to \mu^+\mu^-$. This will be seen in more detail in section 6.6.1.

Figure 6.1 shows the distribution of the variables used in the selection for a sample of reconstructed and MC-truth matched[3] $K^0_S \to \mu^+\mu^-$ and $K^0_S \to \pi^+\pi^-$ events. The vertical line shows the applied cut.

The efficiency of these cuts checked in MC11 samples is $(12.31 \pm 0.07)\%$ for $K^0_S \to \mu^+\mu^-$ and $(19.30 \pm 0.33)\%$ for $K^0_S \to \pi^+\pi^-$, being the p_T cut of 250 MeV/c the dominant inefficiency source. This p_T cut is generically applied to all the *StdParticles* [197]. If just the geometrical cuts are taken into account, the efficiency becomes $(19.12 \pm 0.09)\%$ for $K^0_S \to \mu^+\mu^-$, so compatible with the one of $K^0_S \to \pi^+\pi^-$. To reduce the rate of $K^0_S \to \pi^+\pi^-$, a prescale factor of 1000 is applied in the case of pions, i.e., the selection algorithm for $K^0_S \to \pi^+\pi^-$ is ran only in one each 1000 events.

After the stripping selection, and before the training and application of the multivariate classifier, the following cuts are applied.

- Track momentum smaller than 100 GeV/c (see figure 6.2).
- Secondary vertex χ^2 smaller than 25 (see figure 6.2).
- Lifetime smaller than 130 ps (see figure 6.2).
- Impact parameter of the daughters larger than 0.5 mm (see figure 6.2). This cut has almost no cost on efficiency, and simplifies the trigger studies as the IP cuts applied by the trigger become transparent.
- Λ veto via a cut in the Armenteros–Podolanski [198] plane (see figure 6.3).
- Combined $DLL_{\mu-\pi}$ ($CDLL$) larger than -4, to cut away a large fraction of the double decays in flight with almost no cost in efficiency (see figure 6.4). This cut will be only used for the $K^0_S \to \mu^+\mu^-$ candidates.

The efficiency of this set of cuts on top of the stripping selection is $(98.72 \pm 0.21)\%$ for $K^0_S \to \pi^+\pi^-$ (not applying the PID cuts) and $(98.31 \pm 0.08)\%$ for $K^0_S \to \mu^+\mu^-$ signal, according to MC simulation. If only the geometrical cuts are used, the efficiency for $K^0_S \to \mu^+\mu^-$ becomes $(99.06 \pm 0.06)\%$, also compatible with the one in $K^0_S \to \pi^+\pi^-$.

[3] This refers to the fact that the daughter tracks have most of the hits coming from the correct MC particle.

Variable	$K^0_S \to \mu^+\mu^-$	$K^0_S \to \pi^+\pi^-$
K^0_S mass	[450-1500] MeV/c^2	[400-600] MeV/c^2
K^0_S DOCA	< 0.3 mm	< 0.3 mm
K^0_S DIRA	> 0	> 0
K^0_S τ	$> 0.1 \times \tau(K^0_S)$	$> 0.1 \times \tau(K^0_S)$
K^0_S IP	< 0.4 mm	< 0.4 mm
Daughters IPS	> 10	> 10
Tracks $\chi^2/ndof$	< 5	< 5
Daughters PID	IsMuon=1	-

Table 6.3: Stripping cuts for $K^0_S \to \mu^+\mu^-$ and $K^0_S \to \pi^+\pi^-$. A prescale of 1000 is also applied for the $K^0_S \to \pi^+\pi^-$ case.

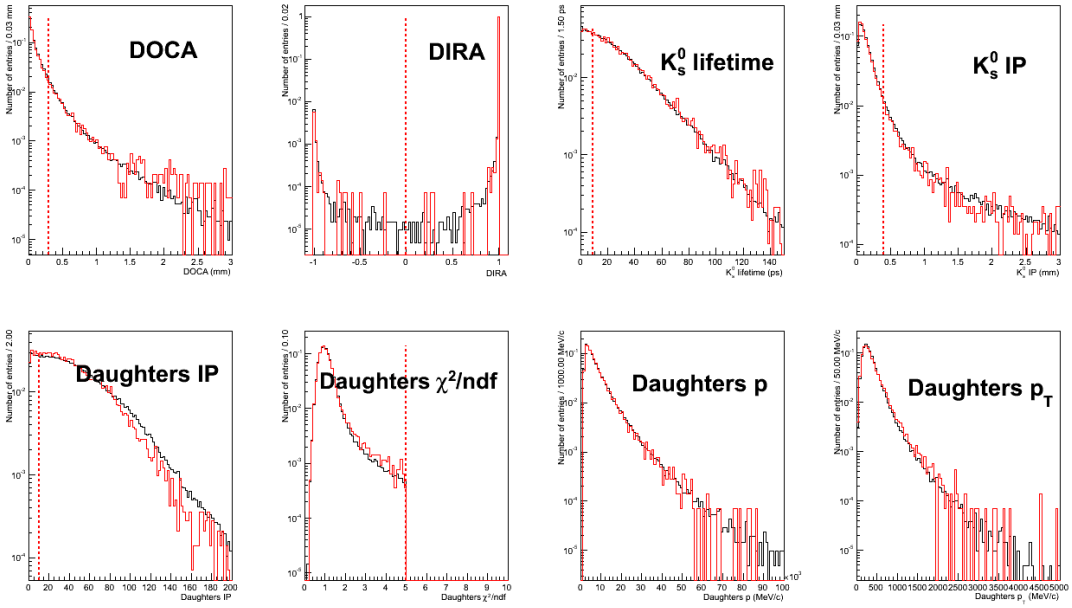

Figure 6.1: Distribution of the variables used in the selection, for $K^0_S \to \mu^+\mu^-$ (black) and $K^0_S \to \pi^+\pi^-$ (red) MC11 sample. The distributions shown were built with MC Truth matched candidates. The vertical line shows the applied cut. Distributions for p and p_T are also shown, since, even if they are not used in the selection, they will be important for normalisation.

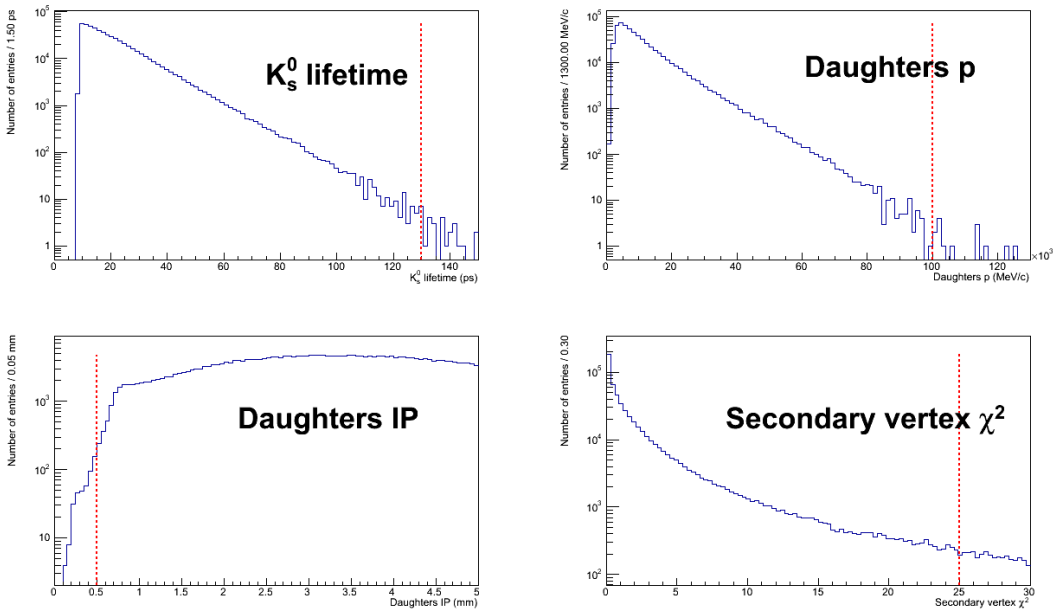

Figure 6.2: Distribution of some of the geometrical variables entering in the fiducial cuts in "minimum bias" $K^0_S \to \pi^+\pi^-$ data. Minimum bias refers to a fraction of the data selected randomly by the trigger, so not biased. The value of the fiducial cuts is indicated by a vertical bar.

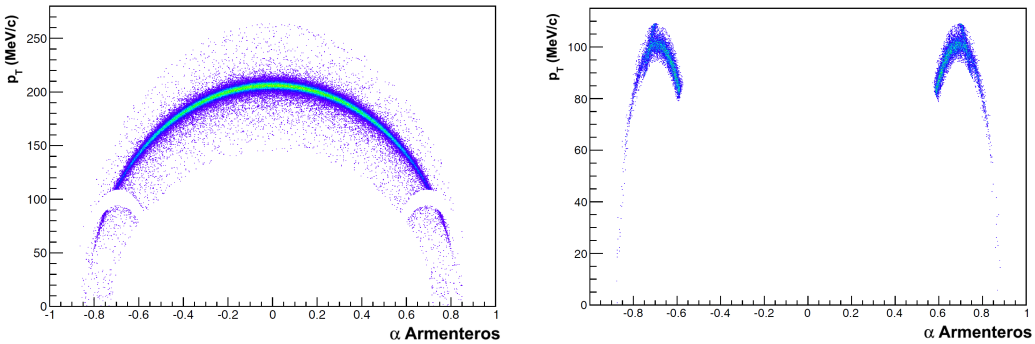

Figure 6.3: Armenteros–Podolanski plot for $K^0_S \to \pi^+\pi^-$ events in data. A cut is applied to remove Λ decays. Left: selected events. Right: rejected events.

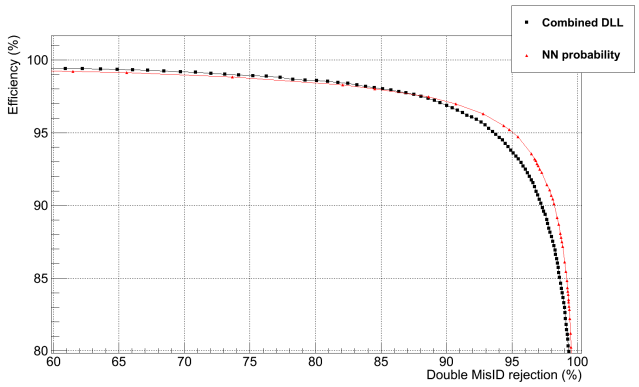

Figure 6.4: Efficiency vs. rejection of doubly misidentified $K^0_S \to \pi^+\pi^-$. The curves are built with cuts in the $CDLL_{\mu-\pi}$ and in the NN_μ [199]. The efficiency (which appears squared) is determined using a data sample of trigger unbiased muons from $B^+ \to J/\psi K^+$, in the p, p_T range of $3-10$ GeV/c and $0.05-1.7$ GeV/c, which mainly corresponds to the one of muons from the signal (see for instance figure 6.1). The misID rejection, from data $K^0_S \to \pi^+\pi^-$ doubly misidentified as $K^0_S \to \mu^+\mu^-$ and also trigger unbiased. This double misID, as it will be seen, is the main component of the left sideband of $K^0_S \to \mu^+\mu^-$. The $K^0_S \to \pi^+\pi^-$ double misID sample has been previously cleaned using a geometrical MVA classifier. The cut finally selected for the analysis, $CDLL > -4$, yields an efficiency of $\sim 98\%$ for $K^0_S \to \mu^+\mu^-$ for a rejection of 80% in double misID $K^0_S \to \pi^+\pi^-$. The efficiency will be carefully obtained in section 6.7.2.

6.4 Trigger aspects

This section summarises the trigger aspects of the $K^0_S \to \mu^+\mu^-$ analysis. This includes identifying the different samples used according to their trigger origin, specifying the lines mainly responsible for the $K^0_S \to \mu^+\mu^-$ trigger and selecting a good proxy for $K^0_S \to \mu^+\mu^-$ as a signal to later train the MVA.

6.4.1 General aspects

A very important difference of this analysis with respect to $B^0_{(s)} \to \mu^+\mu^-$ is the very low trigger efficiency for signal, and the fact that the normalisation channel ($K^0_S \to \pi^+\pi^-$) comes in some cases from a different trigger line.

In particular, the signal and control channels have been separated into the following categories:

- MB (Minimum Bias). LHCb no bias trigger yields a sample of trigger unbiased K^0_S. This sample is negligible in terms of potential $K^0_S \to \mu^+\mu^-$ signal, but it however provides a valuable sample of $K^0_S \to \pi^+\pi^-$ which can be used to understand the distributions that the signal would have in the case of no trigger biases.

- TIS (Trigger Independent of Signal). These events have been triggered by an object different to the signal K^0_S. The trigger biases in this case are small. Also, it is expected that the TIS efficiency is very similar in the signal and the control channel, making quite natural to normalise $K^0_S \to \mu^+\mu^-$ TIS to $K^0_S \to \pi^+\pi^-$ TIS.

- TOS (Trigger On Signal). The muons from $K^0_S \to \mu^+\mu^-$ can fire the LHCb single muon triggers if they have p_T above $\sim 1.3\,\text{GeV}/c$. The summary of the single muon trigger lines, seen in section 3.2.3.1, can be found in table 6.4, with the corresponding global event cuts (GEC) in table 6.5. There is also a small $\sim 7\%$ fraction of TOS coming from charm HLT2 lines, which will be neglected here. The efficiency of this single muon TOS condition is $\sim 1\%$ (according to MC) for those events which would have survived the offline selection. This very low efficiency causes very strong biases in most of the distributions, which become very different to those of the MB events and of TIS events. Also, the branching fraction normalisation in this case is more complicated, as the $K^0_S \to \pi^+\pi^-$ do not fire the muon triggers. This forces the split of the analysis into TIS and TOS categories.

- xTOS (exclusive TOS). Defined as TOS and not–TIS. Most of the TOS will also be xTOS, but this definition excluding the overlap with TIS is needed in order to properly combine the two categories into the final result.

- TOB (Trigger on Both). Defined as triggered but not–TIS and not–TOS. The properties of these events are different to those of the TIS, the TOS, the xTOS and the MB. The analysis explained here does not use $K^0_S \to \mu^+\mu^-$ TOB for the final result.

Line	Cuts
L0 Single Muon	p_T >1.5 GeV/c
Hlt1TrackMuon	p_T >1 GeV/c, IP>0.1 mm, IPχ^2 >16
Hlt2SingleMuon	p_T >1.3 GeV/c, IP>0.5 mm, IPχ^2 >200 Track χ^2 <2. Global Prescale 50 %

Table 6.4: Thresholds of the L0, HLT1 and HLT2 trigger lines relevant for the $K^0_S \to \mu^+\mu^-$ TOS analysis.

trigger level	cut	value
L0	SPD hits	600
HLT1	Velo hits	10000
HLT1	IT hits	3000
HLT1	OT hits	10000 / 15000
HLT2	Velo tracks	350

Table 6.5: Summary of the GECs relevant for $K^0_S \to \mu^+\mu^-$ TOS. As already explained, the number of OT hits upper cut stayed at 15000 for most of the data taking, except for a small period at the beginning when it was set to 10000.

Figures 6.5 and 6.6 illustrate the very different properties of TIS, TOS, and TOB events. For instance, the TOS requirement biases the lifetime distribution towards smaller values. This is due to the fact that, in order to fire the trigger, the K^0_S must have very high p_T. But high p_T K^0_S will decay outside the VELO unless they are very short–lived.

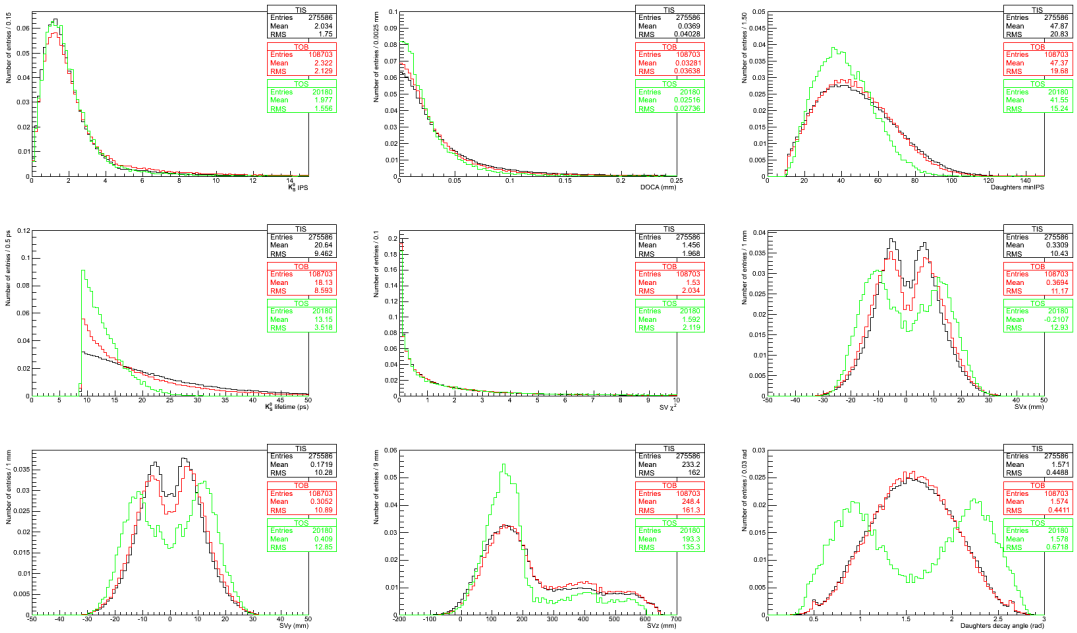

Figure 6.5: Distribution of the variables entering in the MVA computation (see section 6.5) depending on its trigger behaviour (trigger independent of signal, TIS, black line; trigger on signal, TOS, green line; or neither TOS nor TIS, TOB, red line) for $K^0_S \to \pi^+\pi^-$ signal. To reduce any possible background fraction, the signal region has been chosen in the 490-506 MeV/c^2 range.

6.4.2 Defining proxies for $K^0_S \to \mu^+\mu^-$ TOS

While the TIS efficiency is the same for $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu^+\mu^-$, and therefore the $K^0_S \to \pi^+\pi^-$ TIS can perfectly be used both as a normalisation sample and as a training sample for the MVA, this is not the case for TOS events. As the biases in the TOS category are very strong, the $K^0_S \to \pi^+\pi^-$ from MB is also not suitable.

One possible good sample to be considered for training the MVA in the TOS category is $K^0_S \to \pi^+\pi^-$ with double muon misID. However, the trigger efficiency for fake muons is lower

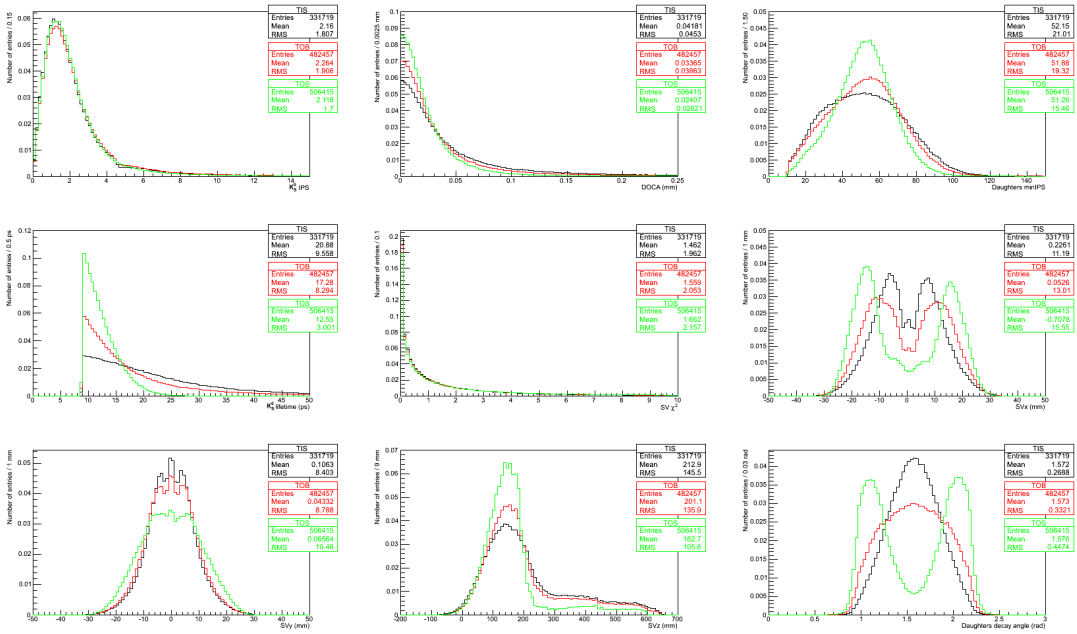

Figure 6.6: Distribution of the variables entering in the MVA computation (see section 6.5) depending on its trigger behaviour (trigger independent of signal, TIS; trigger on signal, TOS; or neither TOS nor TIS, TOB) for $K^0_S \to \mu^+\mu^-$ left sideband (mass below 485 MeV/c^2), corresponding mainly to $K^0_S \to \pi^+\pi^-$ suffering double muon misID.

than for real muons. In order to deal with this, a study has been made to search for a set of muonID cuts that allow the fake muons to have properties closer to the real ones. Choosing a wide range of different PID cuts, the TOS efficiency has been analysed in a set of simulated muons from $K^0_S \to \mu^+\mu^-$ (MC11 generation) and of misidentified pions from $K^0_S \to \pi^+\pi^-$ (MC10 generation). The use of different MC generations comes from the low statistics of the MC11 minbias sample[4], in comparison with the 70 M available in MC10. A possible difference in the performance of the muon trigger in MC10 and MC11 has been checked in $B^+ \to J/\psi K^+$ coming from both generations and selected with the same cuts. As for the trigger configurations used, both belong to 2011. In MC11 the default trigger written corresponds to TCK 0x40760037, while in MC10 the trigger was rewritten with TCK 0x006d0032. The relevant trigger lines for $K^0_S \to \mu^+\mu^-$ include the same cuts in both TCKs. Figure 6.7 shows the TOS efficiency vs. p_T and IP in both samples. They are compatible with being the same, with some small effect coming from the expected different efficiency of the GECs, due the different μ that both samples were generated with.

In order to find the searched correct set of PID cuts, the single muon TOS efficiency has been compared in muons from $K^0_S \to \mu^+\mu^-$ and misidentified pions from $K^0_S \to \pi^+\pi^-$. As explained, the $K^0_S \to \mu^+\mu^-$ belongs to the MC11 generation, while the $K^0_S \to \pi^+\pi^-$ to the MC10. The trigger in the $K^0_S \to \pi^+\pi^-$ has been rewritten using a 2011 TCK (0x006d0032). The PID cuts tuned were:

- Combined DLL (CDLL)
- *nShared* [200]
- $\chi^2/ndof$ of the track

[4]see section 3.2.3.4 for a explanation on MC minbias.

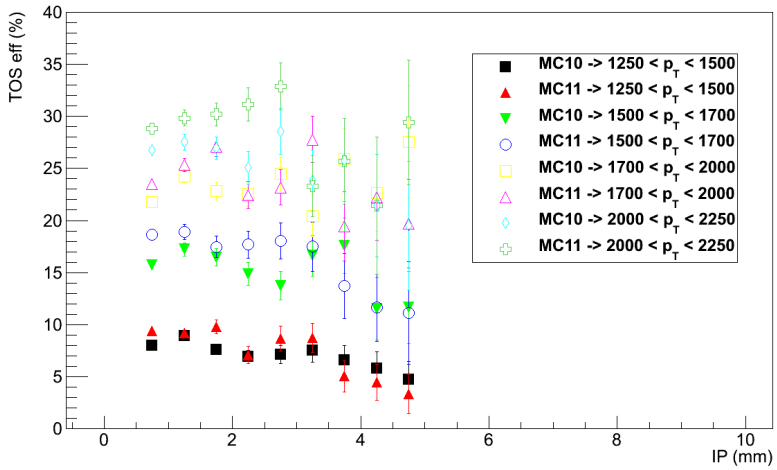

Figure 6.7: Muon TOS efficiency in MC11 and MC10 $B^+ \to J/\psi K^+$ vs. IP and p_T. Both samples were obtained with the same selection and the trigger configurations used include the same cuts in the muon trigger. The p_T bins are defined in MeV/c.

The use of *nShared* can bias the multiplicity of the event. However, as isolation variables will not be used, contrary to the $B^0_{(s)} \to \mu^+\mu^-$ analysis, the *nShared* cuts do not prevent the use of $K^0_S \to \pi^+\pi^-$ misID to train the MVA.

With the idea of building a significant figure of merit, the variable Δ is defined as the difference between the TOS efficiencies in both samples: $\Delta = \left|\epsilon^{SingleMuonTos}_{MC\ K^0_S\to\mu^+\mu^-} - \epsilon^{SingleMuonTos}_{MC\ K^0_S\to\pi^+\pi^-}\right|$. This way, each set of PID cuts will have an associated Δ. Figure 6.8 shows the Single Muon TOS efficiency vs. $1/\Delta$ for real muons from MC11 $K^0_S \to \mu^+\mu^-$, misidentified pions from MC10 $K^0_S \to \pi^+\pi^-$ and misidentified pions from real data. In the three cases, the candidates fulfill the cuts in the stripping selection, including IsMuon in the case of the analysed particle (pion or muon). In the case of the misidentified pions from data, they are selected in the left hand sideband of the $K^0_S \to \mu^+\mu^-$ selection, where they are dominant. For them, the TOS efficiency is estimated as $N^{TOS\&\&TIS}/N^{TIS}$, following the TISTOS method. The points with low values of $1/\Delta$ are related to large differences between the TOS efficiency in muons and misidentified pions. Therefore they correspond to soft cuts in PID. In opposition, for points with larger $1/\Delta$ the trigger behaves similarly for real muons and misID . These correspond to tighter cuts in PID. While the TOS efficiency in muons remains ∼constant for the different cuts in PID (different values of Δ), the misID efficiency increases until it reaches a *plateau*. This is observed as well in the data misID. The selected PID *proxy* cut, with a $1/\Delta$ of ∼30, is $\boldsymbol{nShared < 2\&\&CDLL > -1\&\&\chi^2/ndof < 2.5}$. It is located at the beginning of the *plateau* in figure 6.8. The relevant numbers are summarised in table 6.6.

$\epsilon^{SingleMuonTos}_{MC\ K^0_S\to\mu^+\mu^-}$	$(0.32 \pm 0.02)\%$
$\epsilon^{SingleMuonTos}_{MC\ K^0_S\to\pi^+\pi^-}$	$(0.29 \pm 0.05)\%$
$\epsilon^{SingleMuonTos}_{Data\ K^0_S\to\pi^+\pi^-}$	$(0.82 \pm 0.03)\%$

Table 6.6: Selected PID proxy cut main numbers. The specific cuts are *nShared* <2&&CDLL>-1&&$\chi^2/ndof < 2.5$. The associated $1/\Delta$ is ∼30. See text for details.

The idea to use $K^0_S \to \pi^+\pi^-$ double misID as a proxy for $K^0_S \to \mu^+\mu^-$ in data, in order to train the MVA variable, implies that not only must the trigger behaviour be similar between both but also its geometrical properties remain alike. The geometrical properties of the decays

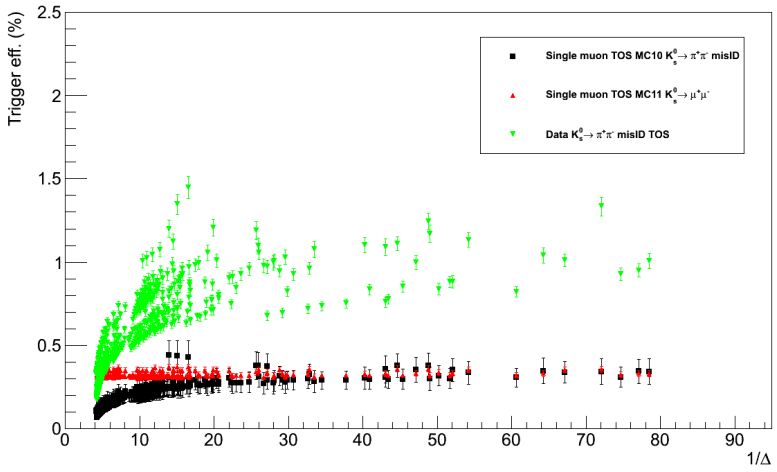

Figure 6.8: Single Muon TOS efficiency vs. $1/\Delta$, defined as $\Delta = |\epsilon^{SingleMuonTos}_{MC\ K^0_S \to \mu^+\mu^-} - \epsilon^{SingleMuonTos}_{MC\ K^0_S \to \pi^+\pi^-}|$. Each point corresponds to a different cut in Combined DLL, $\chi^2/ndof$ and *nShared*. Green points represent misID in data, while black misID in MC10 and red real muons in MC11.

in flight will be very dependent on the *type* of misID, i.e., whether the misidentification was really caused by the decay in flight of the pion, by any other muon entering in its muon system field of interest or even by a punch through. Within the decays in flight, the matching of the tracks with the pion and daughter muon will be also relevant for the geometrical properties. Following all this, the misID (defined as a pion fulfilling IsMuon) is classified in four categories:

1. Track fully matched to the pion, with no decay in flight. It will be dominated by cases in which some muon from the underlying event enters in the field of interest of the pion.

2. Track fully matched to the pion, with decay in flight of the pion. This case is related to a late decay in flight of the pion, mainly taking place at the calorimeter.

3. Track fully matched to the muon. This case is related to a very early decay in flight of the pion, so it mainly happens for low momentum pions.

4. Velo segment of the track matched to pion, Tstations segment of the track matched to the muon. In this case the decay in flight takes place mainly in the magnet region.

In principle, the misID type more interesting for the study is 2, in which on the one hand the track has the same geometrical properties as a muon, and the other, after the proper PID cuts, it performs also similarly at the trigger level. Figures 6.9 and 6.10 show the fraction of misID types after different PID cuts. The plot has been built with misID pions from selected MC10 $K^0_S \to \pi^+\pi^-$. In particular figure 6.9 focuses in cuts in *nShared* and CDLL, while 6.10 in $\chi^2/ndof$. In both figures it can be seen that the cuts in *nShared* mainly reduce the fraction of category 1, while cuts in CDLL, categories 1 and 2. Finally, cuts in $\chi^2/ndof$ reduce the fraction of category 4. From the plots it can be also concluded that, after the proxy selection cut obtained before (*nShared* <2&&CDLL>-1&&$\chi^2/ndof < 2.5$), the remaining sample is dominated by the categories 2 and 4. A softer cut in CDLL or a tighter cut in $\chi^2/ndof$ could enrich the fraction of category 2. However, the first option would decrease the similarity with the trigger for real muons, while the second, since the $\chi^2/ndof$ has a large correlation with the geometrical variables, would also imply a larger difference with the $K^0_S \to \mu^+\mu^-$ sample.

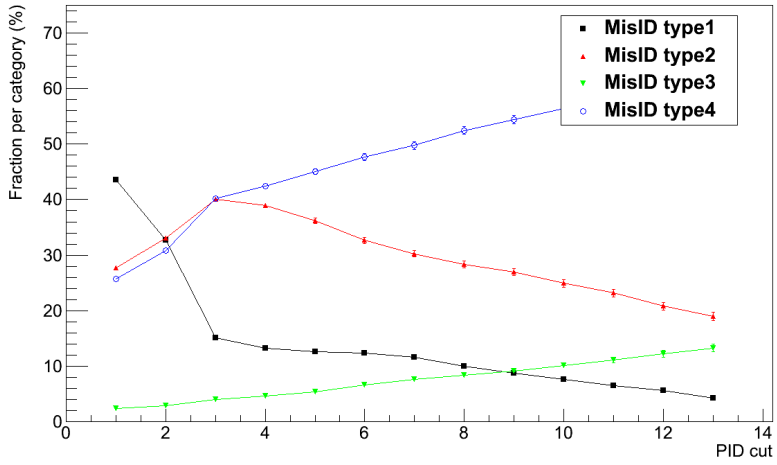

Figure 6.9: Fraction of each misID category vs. different cuts in PID. Cut number 1 is only IsMuon, cut number 2 IsMuon&&$nShared < 2$ and cut number 3 to IsMuon&&$nShared < 2$&&CDLL>-5. Next cuts, from 4 on, are the same than 3 with the CDLL cut increased one unit, up to CDLL>5. For the meaning of the misID categories, see text.

6.4.3 Variables related to the TIS efficiency

The probability to trigger independently of the signal depends mainly on the following variables:

- p_T of the K^0_S. The bigger the p_T of the K^0_S, the higher the TIS efficiency. This comes from the correlation of the K^0_S with the rest of particles produced in the same p–p vertex: the higher the K^0_S p_T the higher the probability to have another particle produced with p_T large enough to fire the trigger.

- Multiplicity of the event. In principle, the higher the multiplicity, the higher should be the probability to have an object (different from the signal K^0_S) that fires the trigger. However, due to the GECs, high multiplicity events are vetoed so that the final dependency on the multiplicity is not trivial.

- K^0_S Impact parameter. Since the K^0_S have a non negligible probability to come from detached objects (bottom, charm, hyperons), the sisters of the K^0_S can also make tracks with high impact parameter and increase their chances to trigger. Indeed, a small correlation between the IPS of the K^0_S and the TIS efficiency is found.

Figure 6.11 shows the dependency of the TIS efficiency with the p_T of the K^0_S, the multiplicity (defined here as the number of *long tracks*) and the IPS of the K^0_S.

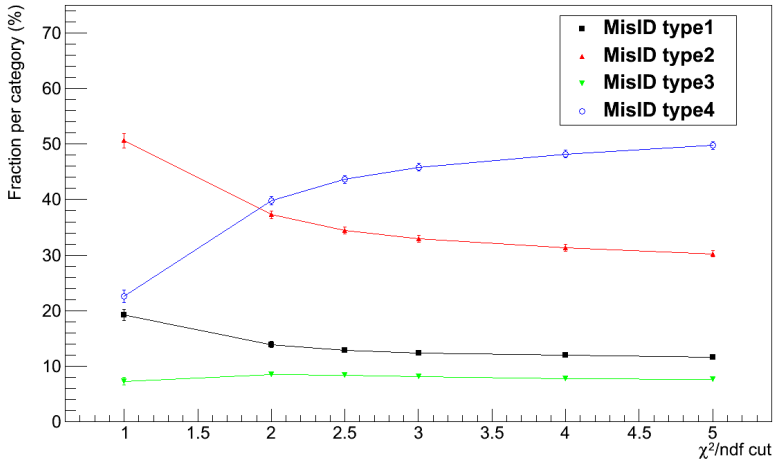

Figure 6.10: Fraction of each misID category vs. different upper cuts in $\chi^2/ndof$. An extra cut of IsMuon&&$nShared < 2$&&CDLL>-1 is also applied in all the cases. This way, the fraction seen with the cut in $\chi^2/ndof < 2.5$ corresponds to the fraction after the PID proxy cut. For the meaning of the misID categories, see text.

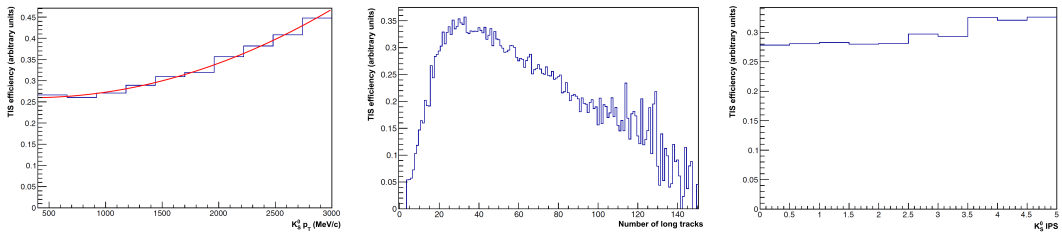

Figure 6.11: Dependency of the TIS efficiency with: K^0_S p_T (left, the red line being a parabolic fit), number of long tracks (center), and K^0_S IPS (right). The y axis is in arbitrary units.

6.5 Multivariate classifier

In order to increase background rejection, a *Boosted Decision Tree* (BDT) is used. Note that the BDT was also the MVA chosen for the $B^0_{(s)} \to \mu^+\mu^-$ analysis, as seen in section 5.7. The variables that allow the separation between signal and background have very different distributions in TIS and TOS events (see figures 6.5 and 6.6). This forces the separation of the BDT training into two, having a BDT for TOS and another for TIS. Also, the nature of the background is different in TIS and TOS events and, in particular, the fraction of background from material interactions is bigger in the TOS category due to its harder p_T spectrum, as discussed in what follows.

6.5.1 Choice of variables. Background from material interaction

The following variables are chosen, as a minimal set to describe the topology and kinematics of the signal decay.

- *Lifetime of the K^0_S candidate.* This variable is computed using the distance between the reconstructed Secondary (SV) and Primary (PV) vertices, and the reconstructed momentum of the K^0_S candidate. When more than one PV is reconstructed, the one that gives the minimum K^0_S impact parameter significance is chosen.

- *Muon impact parameter significance.* This variable is the lowest impact parameter significance of the two muon candidates with respect to any of the primary vertices reconstructed in the event.

- K^0_S *impact parameter significance.*

- *DOCA.* This variable is the distance of closest approach between the two muon candidates.

- *Secondary vertex* χ^2. This variable adds complementary information with respect to DOCA, as it uses information on the error of the vertex fit. However, it does not substitute DOCA, as the χ^2 can be good for vertices that are poorly reconstructed.

- *Decay angle of the daughters.* It has to be isotropic for signal, but not necessarily for background.

It has also been found that an important amount of background comes from material interactions (see figure 6.12). To deal with this, two different approaches are tried:

- *Coordinates of the secondary vertex in the laboratory rest frame.* While the lifetime explodes the large distance of flight of the K^0_S mesons, the position of the secondary vertex is included because an important fraction of the background comes from material interactions (see figure 6.12). In this case, the p_T of the K^0_S cannot be included because the simultaneous use of the lifetime, the p_T, and the SV position allows the MVA to learn which is the mass of the candidate, and thus using it to discriminate. In such a case, the background rejection would be very different in the sidebands and in the signal region, and a fake signal peak could be artificially created. This is the approach that works better for the events in the TIS category. Alternatively, the p_T of the K^0_S could have been included if the lifetime was dropped.

- *Boolean variable IsMaterial.* This variable is defined in [201] to know whether a given decay vertex coincides with a point in the material or not. In this way, the p_T of the K^0_S can also be used as a variable, and one does not rely on the ability of the BDT to use the SV position to know whether it is in the material or not. This second approach is the best for the TOS category, where the fraction of events from material interactions is found to be much bigger than in the TIS case. This seems to be caused by the fact that the muons from the interactions have harder p_T spectrum than those from other background sources. This is shown in figure 6.13, where the p_T spectrum of TIS $K^0_S \to \mu^+\mu^-$ background is compared between the events with $IsMaterial = 1$ and $IsMaterial = 0$.

The variables included in the TIS and TOS MVA are summarised in table 6.7.

TIS	TOS
Lifetime of the K^0_S candidate	Lifetime of the K^0_S candidate
Muon impact parameter significance	Muon impact parameter significance
K^0_S impact parameter significance	K^0_S impact parameter significance
DOCA	DOCA
Secondary vertex χ^2	Secondary vertex χ^2
Decay angle of the daughters	Decay angle of the daughters
3D Coordinates of the secondary vertex	Boolean variable *IsMaterial*
	K^0_S p_T

Table 6.7: Summary of variables entering in the TIS and TOS MVA computation

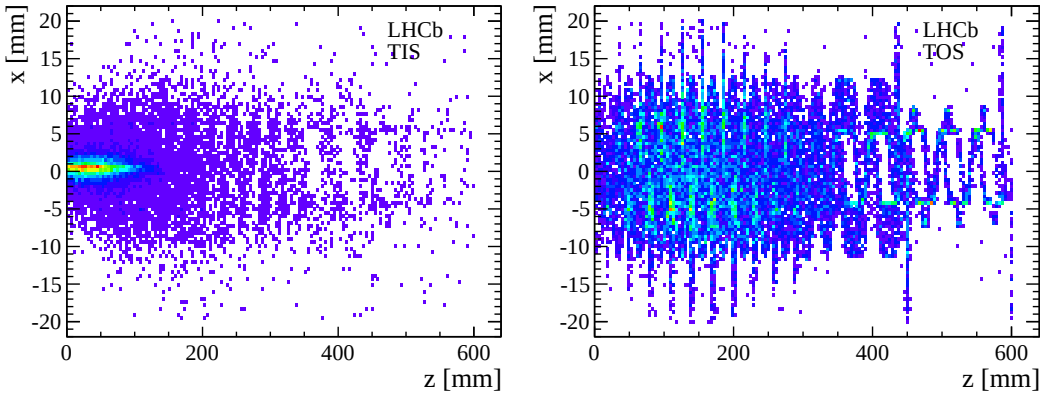

Figure 6.12: x vs. z position of the secondary vertices of the background candidates found in the data mass right-hand sideband in the TIS (left) and TOS (right) categories. The plots show that a sizeable amount of the background comes from material interactions in the region of the VELO. This is even more important in the TOS case.

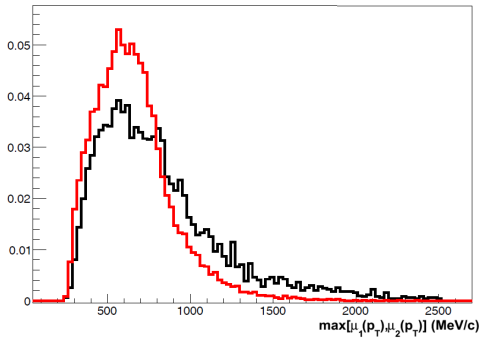

Figure 6.13: Distribution of the maximum p_T of the muon candidates in $K^0_S \to \mu^+\mu^-$ TIS events from the right–hand sideband. Red: events with *IsMaterial* = 0. Black: events with *IsMaterial* = 1. The events with *IsMaterial* = 1 have harder p_T spectrum, increasing their probability to fire the single muon triggers.

6.5.2 Training and expected performance

As there is no large sample of background suitable for $K^0_S \to \mu^+\mu^-$ studies, the MVA has to be trained on data. As a signal proxy, $K^0_S \to \pi^+\pi^-$ TIS is used for the $K^0_S \to \mu^+\mu^-$ TIS category and $K^0_S \to \pi^+\pi^-$ double misID TOS, with the cuts seen in section 6.4.2, for the $K^0_S \to \mu^+\mu^-$ TOS category. For the background, the corresponding TIS and TOS $K^0_S \to \mu^+\mu^-$ sidebands in the region below 1 GeV/c^2 are chosen. To avoid biasing the background estimation (as the same sidebands are going to be used for background interpolation), the data is split in two samples of approximately the same size:

- Sample A: events with run number smaller than 97500.
- Sample B: events with run number bigger than 97500.

The MVA is trained in sample A and applied in sample B. Then, another MVA operator is trained in sample B to later be applied in sample A. In this way, the background estimation cannot be biased by any overtraining. The $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ measurement will consist in a combination of the results in sample A and sample B, as well as of TOS and TIS, i.e, a total of 4 independent data samples.

In summary, the training samples of the BDT are the following:

- For TIS events: $K^0_S \to \pi^+\pi^-$ TIS as signal (neglecting the background present under the mass peak) and the upper mass sideband of TIS events as background. The region above 1.1 GeV/c^2 is used to decide the BDT tuning, and once the tuning is decided, the region below 1.0 GeV/c^2 is used to train the final BDT.

- For TOS events: $K^0_S \to \pi^+\pi^-$ misID TOS (with the cuts on the muons: *nShared* <2, CDLL>-1, track $\chi^2/ndof < 2.5$, which is the proxy cut defined in section 6.4.2) as signal and the upper mass sideband of TOS events as background. The region above 1.1 GeV/c^2 is again used to decide the BDT tuning, and once the tuning is decided, the region below 1.0 GeV/c^2 is used to train the final BDT.

To choose a particular MVA algorithm, the $K^0_S \to \pi^+\pi^-$ signal and the $K^0_S \to \mu^+\mu^-$ mass sidebands above 1 GeV/c^2 are used, thus avoiding any overlap with the final interpolation sample. Several of the usual TMVA tunings for BDT and NN have been tested. The BDTD was found to be the best choice. NeuroBayes NN was also tested, without any special tuning, but was found to be slightly worse than the BDT one. Further studies on the tuning of these methods could improve the performance of the MVA algorithms.

Figure 6.14 (figure 6.15) shows the ROC curve for different TMVA BDT tunings in the TIS (TOS) case. The BDT distributions on data signal and background can be seen in figure 6.16 for the TIS and TOS cases.

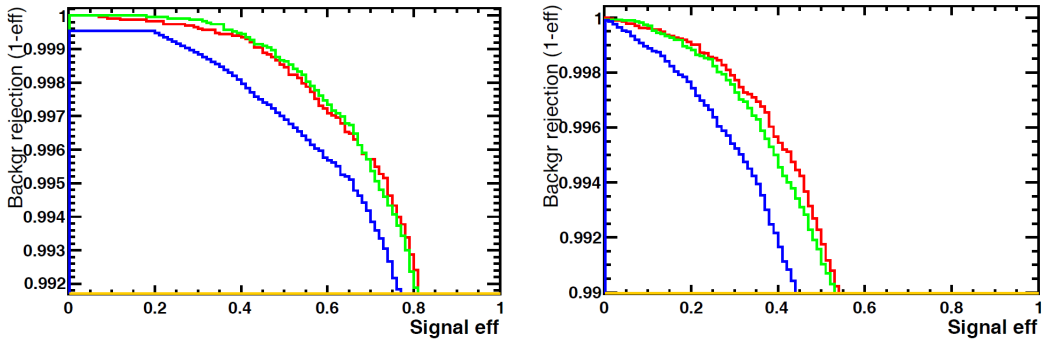

Figure 6.14: Performance of the Boosted Decision Trees tested in the TIS category. In all figures the y axis corresponds to background rejection and the x axis to signal efficiency. The different colours correspond to different (TMVA default) BDT tunings: BDT (red), BDTG(blue), BDTD(green). Left: performance in the samples used to decide the best BDT: $K^0_S \to \pi^+\pi^-$ as signal proxy, and $K^0_S \to \mu^+\mu^-$ mass sidebands above 1.1 GeV/c^2 as background proxy. Right: Performance using as a background the right sideband up to 1.0 GeV/c^2. This second plot is only illustrative, and has not been used to decide the BDT.

6.5.3 BDT lineshape. p_T cut for $K^0_S \to \pi^+\pi^-$

The big differences between TOS events and both TIS and MB events can make the normalisation to $K^0_S \to \pi^+\pi^-$ very difficult. In particular, a $\sim$ 50% efficient BDT cut for $K^0_S \to \mu^+\mu^-$ TOS events becomes only $\sim$ 5% efficient for $K^0_S \to \pi^+\pi^-$ from the minimum bias, i.e, they differ by an order of magnitude. Because of this, to make the phase space of the signal and normalisation channels closer, a cut of 1.3 GeV/c on the maximum p_T of the daughters is applied for the TOS analysis. Using this cut, the BDT of $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu^+\mu^-$ TOS becomes much more similar, as it will be seen in figure 6.18.

Another important remark is that events with values of BDT smaller than 0.2 are not considered because of the large amount of background at low BDT values (see figure 6.16). The rest of the events are classified in 10 bins of $\sim$ 10% signal efficiency each.

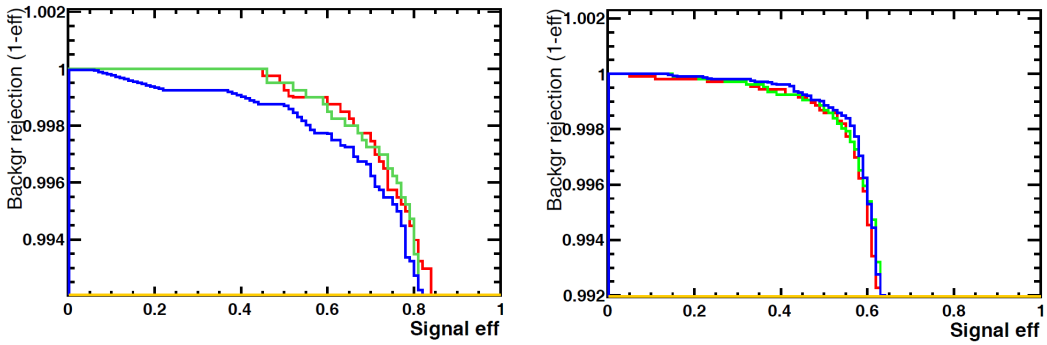

Figure 6.15: Performance of the Boosted Decision Trees tested in the TOS category. The y axis corresponds to background rejection and the x axis to signal efficiency. The different colours correspond to different (TMVA default) BDT tunings: BDT (red), BDTG(blue), BDTD(green). The samples used to decide the best BDT: $K^0_S \to \pi^+\pi^-$ misID (using the muon cuts specified in the text) as signal proxy, and $K^0_S \to \mu^+\mu^-$ mass sidebands above 1.1 GeV/c^2 as background proxy. Left: performance in the samples used to decide the best BDT. Right: Performance using as a background the right sideband up to 1.0 GeV/c^2. This second plot is only illustrative, and has not been used to decide the BDT.

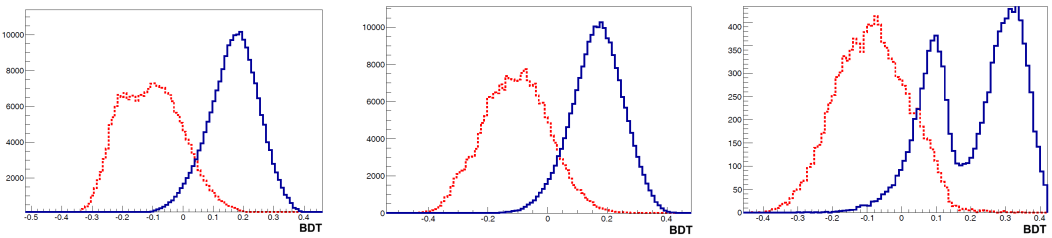

Figure 6.16: Distribution of the BDT response. Red: $K^0_S \to \mu^+\mu^-$ mass sidebands. Blue: $K^0_S \to \pi^+\pi^-$ events. Left: sample A TIS. Center: sample B TIS. Right: sample B TOS. The double peak of the TOS signal reflects the boolean behaviour of *IsMaterial*

The BDT bins are defined differently for TOS and TIS categories, but numerically the same for samples A and B. This is done to simplify the analysis, even if it may produce non-flat BDT efficiencies per bin in the samples in which the bins were not defined. The bins for BDT TIS are built using $K^0_S \to \pi^+\pi^-$ TIS sample B, while the bins for BDT TOS using $K^0_S \to \pi^+\pi^-$ double misID TOS (from left hand sideband of $K^0_S \to \mu^+\mu^-$) after applying the PID cuts seen in section 6.4.2 and also used for training the signal. With this, the efficiencies per BDT bin for sample A are non-flat. This does not imply a loss or gain in sensitivity, since it coherently affects the background. The efficiency per BDT bin in TIS and TOS categories for signal and background in samples A and B can be seen in figure 6.17. Here, the signal sample is the same than used for defining the bins (so that the efficiency is flat for sample B), while for background the right hand sideband of $K^0_S \to \mu^+\mu^-$ (in the 600-1000 MeV/c^2 range) has been chosen, after applying the selection and fiducial cuts (including PID). It must be said that most of the background has BDTs below 0.2, as seen in figure 6.16, and the distributions seen in the figure are normalised to the total amount of candidates with BDT above 0.2.

The efficiency as a function of the BDT cut is also shown in figure 6.18, comparing in this case data and MC samples. The figure shows two distinctive features: the approximate similarity between $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu^+\mu^-$ BDT lineshapes, but also the similarity between data and MC distributions, even after hard p_T cuts.

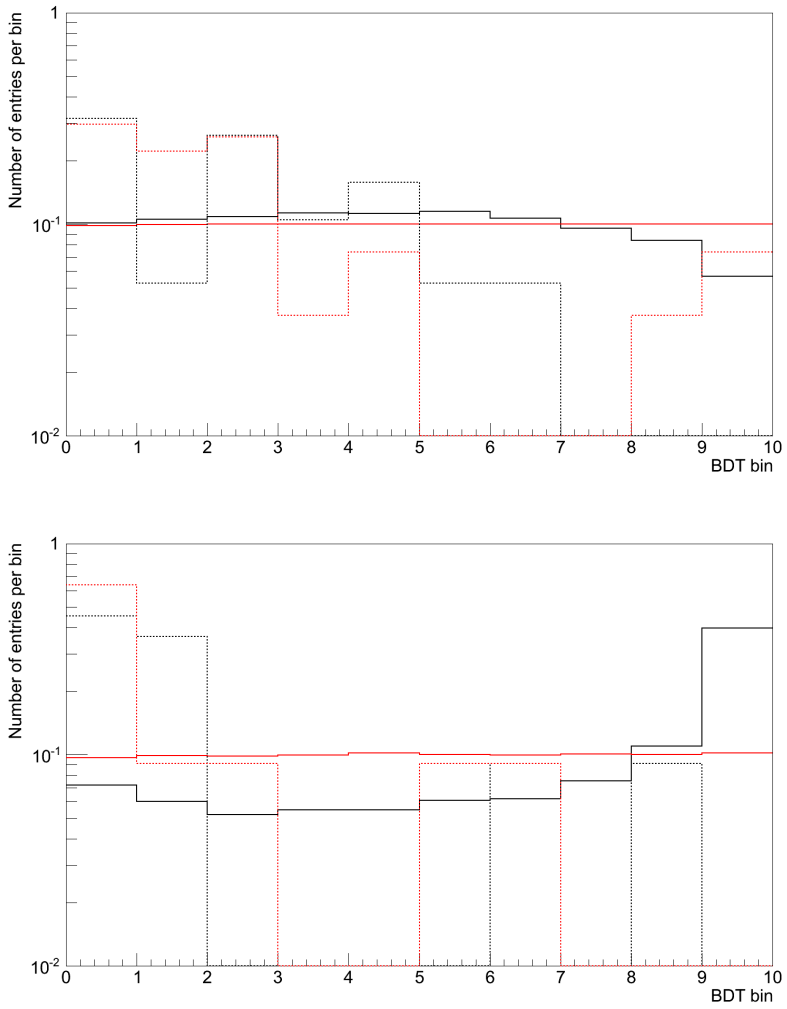

Figure 6.17: Efficiency per BDT bin in TIS (top) and TOS (bottom) categories for signal and background in samples A and B. Black lines correspond to sample A, while red lines to sample B. The straight lines are built with signal samples, while dashed lines with background samples. The signal sample for the TIS category is $K^0_S \to \pi^+\pi^-$ TIS, while for the TOS category $K^0_S \to \pi^+\pi^-$ double misID TOS (from left hand sideband of $K^0_S \to \mu^+\mu^-$) after applying the PID cuts seen in section 6.4.2. The background sample is built with the right hand sideband of $K^0_S \to \mu^+\mu^-$ (in the 600-1000 MeV/c^2 range) after applying the selection and fiducial cuts (including PID). The distributions seen in the figure are normalised to the total amount of candidates with BDT above 0.2.

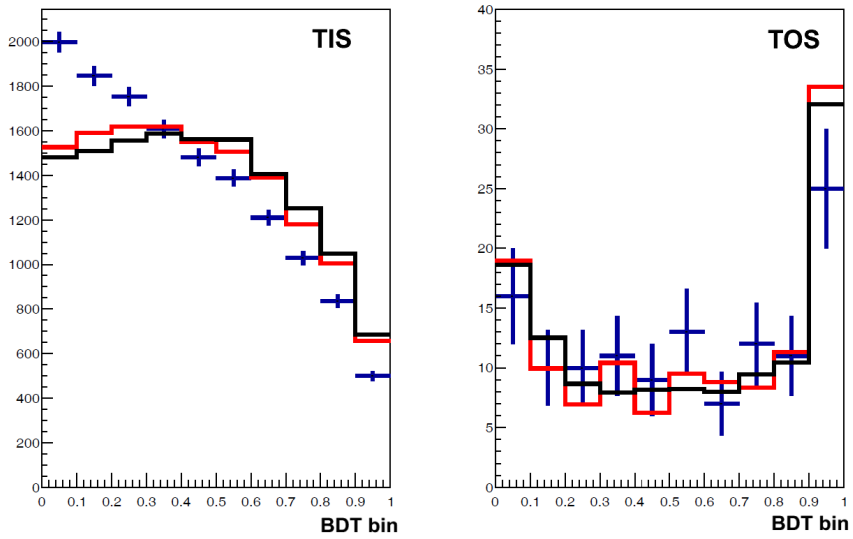

Figure 6.18: Efficiency per BDT bin for $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu^+\mu^-$. Blue: $K^0_S \to \mu^+\mu^-$ MC, red: $K^0_S \to \pi^+\pi^-$ MC, black: $K^0_S \to \pi^+\pi^-$ MB data. Left: BDT for TIS (note that the samples in the plot are minimum bias, no TIS requirement applied). Right: BDT for TOS, including the 1.3 GeV/c cut on the maximum p_T of the daughters. The $K^0_S \to \mu^+\mu^-$ are required to be TOS in the right hand plot.

6.6 Specific backgrounds

The specific backgrounds (other than combinatorial) for $K^0_S \to \mu^+\mu^-$ will be studied in this section. These include $K^0_L \to \mu^+\mu^-$, $K^0_S \to \pi^+\pi^-$ in which both pions decay to a neutrino and a muon, $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ and $K^0_S \to \mu^+\mu^-\gamma$. The first, which could peak in the $K^0_S \to \mu^+\mu^-$ signal region, will be shown to be negligible for the current dataset. The second will be the main component of the left sideband of $K^0_S \to \mu^+\mu^-$, with a tail that enters in the $K^0_S \to \mu^+\mu^-$ signal region. The best parameterisation for this tail will be studied with the help of MC simulation. Finally, the third and fourth, which are expected to lie in the left sideband, will also be seen to be unharmful here.

6.6.1 $K^0_L \to \mu^+\mu^-$

$K^0_L \to \mu^+\mu^-$ events are topologically equivalent to $K^0_S \to \mu^+\mu^-$ if the K^0_L decays early enough and the muons make long tracks. The $K^0_L \to \mu^+\mu^-$ decays can be disentangled from $K^0_S \to \mu^+\mu^-$ only by the lifetime (or variables which depend on it). The $\mathcal{B}(K^0_L \to \mu^+\mu^-)$ is $(6.84\pm0.11)\times10^{-9}$ [5], much larger than the SM expectation of $\mathcal{B}(K^0_S \to \mu^+\mu^-)$, $(5.0 \pm 1.5) \times 10^{-12}$.

In order to get the $K^0_L \to \mu^+\mu^-$ yield, the ratio of efficiencies for $K^0_S \to \mu^+\mu^-$ and $K^0_L \to \mu^+\mu^-$ is estimated. This ratio is assumed to depend only on the lifetimes of K^0_S and K^0_L (so neglecting resolution effects), through the non-flat lifetime acceptance ($Acc[t]$):

$$\frac{\epsilon_{K^0_L\to\mu^+\mu^-}}{\epsilon_{K^0_S\to\mu^+\mu^-}} = \frac{\dfrac{\displaystyle\int_0^\infty Acc(t)e^{-\Gamma_L t}dt}{\displaystyle\int_0^\infty e^{-\Gamma_L t}dt}}{\dfrac{\displaystyle\int_0^\infty Acc(t)e^{-\Gamma_S t}dt}{\displaystyle\int_0^\infty e^{-\Gamma_S t}dt}} \tag{6.1}$$

where $1/\Gamma_S$ ($1/\Gamma_L$) is the mean lifetime of K^0_S (K^0_L).

In order to obtain the functional form of $Acc(t)$, the lifetime distribution of the $K^0_S \to \pi^+\pi^-$ TIS events in data is used. As the intention here is to obtain an estimation of the order of magnitude of this background, the description of the of the lifetime acceptance will only be approximate, without requiring a high degree of precision. This description is sufficient as it will be seen that the contribution of this background is negligible.

In the lifetime range accepted by the selection and fiducial cuts (8.95 to 130 ps), it can be seen that the distribution in data behaves approximately like an exponential (see figure 6.19):

$$p.d.f.(t) \propto e^{\alpha t} \tag{6.2}$$

Thus, the acceptance in that range is, to the degree of precision that needed for this study, also exponential:

$$p.d.f.(t) \propto e^{\alpha t} = Acc(t)e^{-\Gamma_S t} \to Acc(t) = e^{(\alpha+\Gamma_S)t} = e^{\beta t} \tag{6.3}$$

The lifetime distribution of $K^0_S \to \pi^+\pi^-$ TIS events in data is fitted to the expression in equation 6.2. The fit is done separately in events from sample A and sample B (see figure 6.19). The results of the fit obtained in each sample are numerically similar (within $\sim 1.5\%$) , although not compatible within the errors provided by the fit. This can be due to small differences in the running conditions or that the exponential model is not accurate enough for large statistics. However, the model is good enough for the purpose of estimating the $K^0_L \to \mu^+\mu^-$ background, as the conclusions will not depend on a precise value of α. At the end of this section the lifetime acceptances in the TOS category will be briefly discussed.

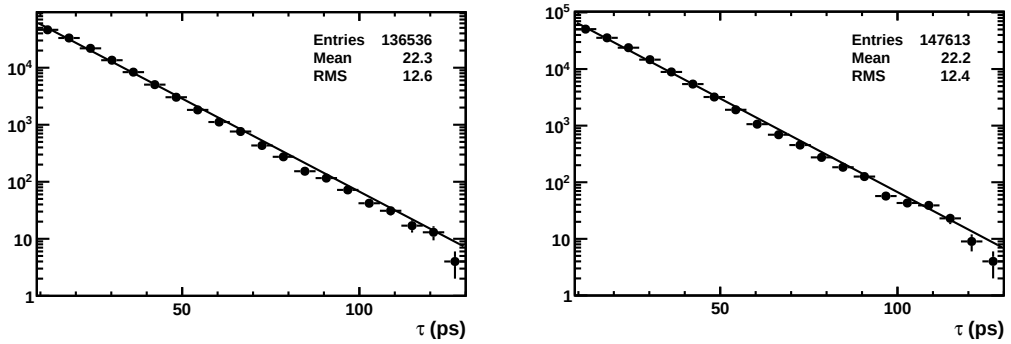

Figure 6.19: Lifetime distributions observed in $K^0_S \to \pi^+\pi^-$ TIS events in data. The line corresponds to a fit to $e^{\alpha t}$. Left: sample A, which yields $\alpha_A = -75.1 \pm 0.2$ ns^{-1}. Right: sample B, which yields $\alpha_B = -76.1 \pm 0.2$ ns^{-1}. The two results are numerically similar, although not compatible within the errors given by the fit.

From equation 6.3 and considering that the acceptance is 0 outside the range accepted by the selection, the following expression is obtained, according to (6.1):

$$\frac{\epsilon_{K^0_L\to\mu^+\mu^-}}{\epsilon_{K^0_S\to\mu^+\mu^-}} = \frac{\Gamma_L}{\Gamma_S}\frac{(\beta-\Gamma_S)}{(\beta-\Gamma_L)}\frac{[e^{(\beta-\Gamma_L)t_1} - e^{(\beta-\Gamma_L)t_0}]}{[e^{(\beta-\Gamma_S)t_1} - e^{(\beta-\Gamma_S)t_0}]} \tag{6.4}$$

being $t_0 = 8.95$ ps and $t_1 = 130$ ps the range allowed by the selection, $\Gamma_S = 1/89.5$ ps^{-1} and $\Gamma_L = 1/5.1 \times 10^{-4}$ ps^{-1}.

Equation 6.4 yields an efficiency ratio of $\sim 2.3 \times 10^{-3}$. The exact value of the efficiency ratio depends on the exact value of β, but even for $\beta = 0$ an efficiency ratio of $\sim 3.5 \times 10^{-3}$ is obtained. The formula ignores minor effects coming from the fact that some K^0_L with lifetimes bigger than 130 ps can be selected because of resolution effects. However, the result of equation 6.4 does not depend significantly on t_1. Thus, the effective branching fraction of $K^0_L \to \mu^+\mu^-$ becomes 1.6×10^{-11}. This background can therefore be neglected for the current analysis.

It should be also remarked that equal production cross sections for K^0_S and K^0_L have been assumed. This assumption has been checked by counting the amount of those particles produced in MC minimum bias and classifying them by the particle ID of the *kaon* mother. Table 6.8 shows the observed K^0_S yields in MC minimum bias as well as the ratio of $\frac{K^0_S}{K^0_L}$, for particles that do not come from material interactions. The total ratio is 1.0182 ± 0.0065. The D^0 produces sizeable different K^0_S and K^0_L yields, but the importance of the D^0 over the total yield, dominated by K^0, is negligible. The interactions with material can also change the ratio of effective cross sections. This is illustrated in table 6.9. However, it has been found that none of the selected MC signal candidates comes from material interactions, so the corrections that such interactions produce to the ratio of K^0_S and K^0_L cross sections are neglected. Figure 6.20 shows the ratio of K^0_S to K^0_L yields in bins of longitudinal momentum and polar angle. It is found that this ratio is quite flat and any correction to the ratio of efficiencies other than the one due to the lifetime is considered.

It is important to notice that, from the obtained effective branching fraction, this background will be important to reach $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ at the 10^{-11} range. Moreover, according to the current study, the exponential decay of the K^0_L will be seen as

$$p.d.f.(t)_{K^0_L} \propto e^{\alpha_L t} = e^{(\beta-\Gamma_L)t} \sim e^{-0.064t(\text{ps})} \tag{6.5}$$

not so different from the distribution of the K^0_S, $\sim e^{-0.075t(\text{ps})}$.

These calculations come from the lifetime acceptance measured in the TIS category. In the TOS category (see figure 6.21), the lifetime acceptance drops much faster, which will lead to a

Particle	$N_{K^0_S}$	$N_{K^0_S}/N_{K^0_L}$
$f_2(1270)$	108	1.13 ± 0.16
D^0	252	1.76 ± 0.18
$a_0(980)^0$	46	1.00 ± 0.21
$\phi(1020)$	1495	1.00 ± 0.037
$f_2'(1270)$	174	0.95 ± 0.10
$a_2(1320)^0$	110	1.15 ± 0.16
D_s^+	11	1.10 ± 0.48
K_0	47140	1.0162 ± 0.0066
$D^\pm$	92	1.10 ± 0.17

Table 6.8: K^0_S yields produced by different particle types in minimum bias MC, as well as the ratio of K^0_S over K^0_L produced by such particles. This table shows the numbers for kaons not coming from material interactions.

Particle	$N_{K^0_S}$	$N_{K^0_S}/N_{K^0_L}$
n	644	0.921 ± 0.052
K^+	1921	1.25 ± 0.04
K^0_L	13237	0.421 ± 0.007
p	430	0.979 ± 0.067
γ	13	1.08 ± 0.41
Σ^-	13	0.46 ± 0.23
Ξ^-	4	0.50 ± 0.43
Λ	86	1.35 ± 0.19
π^+	2197	1.03 ± 0.03
K^0_S	216	3.65 ± 0.28
nucleus	214	4.2 ± 0.32
Σ^+	11	1.27 ± 0.51

Table 6.9: K^0_S yields produced by different particle types in minimum bias MC, as well as the ratio of K^0_S over K^0_L produced by such particles. This table shows the numbers for kaons coming from material interactions.

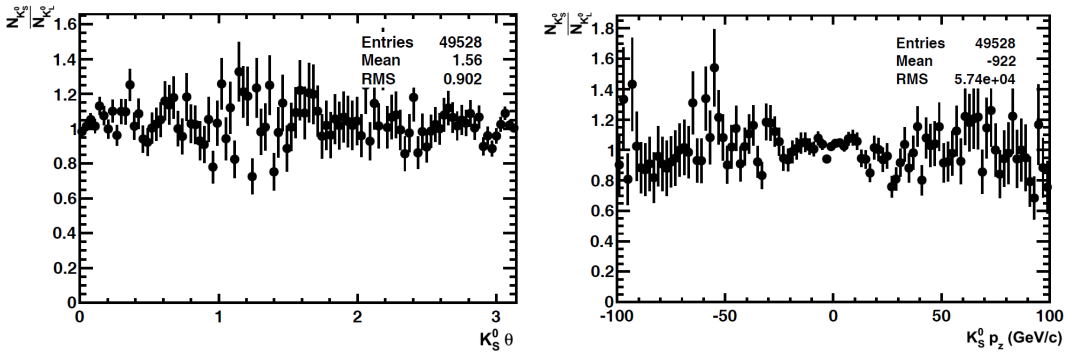

Figure 6.20: $\frac{N_{K^0_S}}{N_{K^0_L}}$ in MC minimum bias for particles produced before material interaction. Left: as a function of the polar angle θ of the kaon. Right: as a function of the p_z of the kaon, in GeV/c.

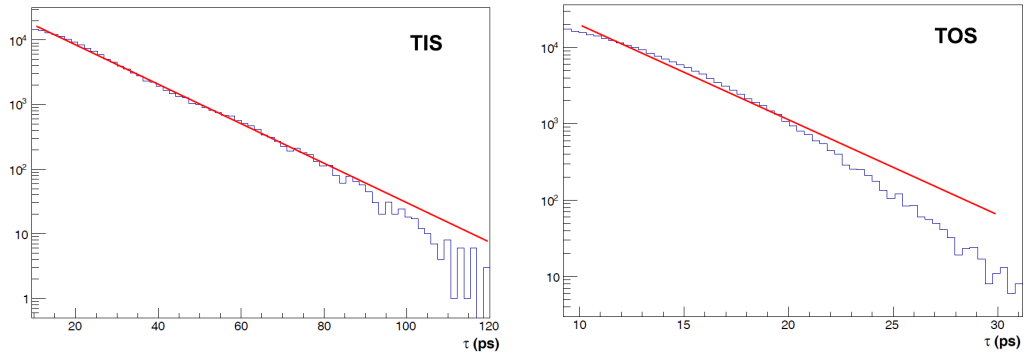

Figure 6.21: Lifetime distributions observed in $K^0_S \to \pi^+\pi^-$ with double misID events in sample A in data. The line corresponds to a fit to $e^{\alpha t}$. Left: TIS events yield $\alpha_{TIS} = -70.3 \pm 0.1\ ns^{-1}$. Right: TOS yield $\alpha_{TOS} \sim -300\ ns^{-1}$, but the distribution clearly differs from an exponential function.

smaller $K^0_L \to \mu^+\mu^-$ effective branching ratio.

Finally, a possible contribution to the background of $K^0_L \to \mu^+\mu^-\gamma$ has also been taken into account, since the $\mathcal{B}(K^0_L \to \mu^+\mu^-\gamma) = 3.6 \times 10^{-7}$ [5] is a factor of ∼50 larger than that of $K^0_L \to \mu^+\mu^-$. Following the result for the effective $\mathcal{B}(K^0_L \to \mu^+\mu^-)$ obtained before and assuming the same lifetime acceptance for $K^0_L \to \mu^+\mu^-$ and $K^0_L \to \mu^+\mu^-\gamma$, it is easy to obtain

$$\mathcal{B}(K^0_L \to \mu^+\mu^-\gamma)|_{eff} = 8.5 \times 10^{-10}$$

It is also important to say that a possible background from $K^0_L \to \mu^+\mu^-\gamma$ would not anyway be expected to peak at the K^0_S mass, because of the energy taken by the non-reconstructed photon. In this way, the reconstructed K^0_S mass would lie in the lower mass sideband, producing because of this a significantly smaller effective $\mathcal{B}$ in the $K^0_S \to \mu^+\mu^-$ selection mass window. With all this, the contribution from $K^0_L \to \mu^+\mu^-\gamma$ should be small when compared to the upper limits expected to be set in the $\mathcal{B}(K^0_S \to \mu^+\mu^-)$, but again to be revisited in future searches.

6.6.2 $K^0_S \to \pi^+\pi^-$ double misID

The $K^0_S \to \pi^+\pi^-$ decays in which both pions are misidentified as muons turn out to be the most relevant specific background for $K^0_S \to \mu^+\mu^-$. Geometrically, the decay is equivalent to

the signal, so the BDT is not effective to reject it. For the muon PID, it typically involves low momentum pions, where the muonID algorithms are less performing (see sections 3.2.3.3.3, 5.5, 5.6.1). However, the excellent LHCb mass resolution makes this background much less harmful, as it is not peaking in the signal region, but a tail entering on it that can be parameterised on data. There are two factors that "move" the $K^0_S \to \pi^+\pi^-$ double misID mass distribution to the left hand sideband of $K^0_S \to \mu^+\mu^-$. First, changing the mass hypothesis from $\pi^+\pi^-$ to $\mu^+\mu^-$, which is more relevant than in the B decays case, due to the difference in mass between muons and pions being not negligible when compared to the K^0_S one. Second, the fact that the measured momentum of the decays in flight tends to be smaller than that of the original pion. Figure 6.22 left shows the mass spectra seen for the $K^0_S \to \mu^+\mu^-$ candidates after the stripping and fiducial cuts, for both the TIS and TOS categories together. The missing part in the centre of the spectra is the signal region, which is blinded. The right hand sideband includes some combinatorial background component, which will be mainly rejected by the BDT cuts. It can be also seen that the dominant component of the selected candidates is the double misID from $K^0_S \to \pi^+\pi^-$, which peaks in the left hand sideband. Figure 6.22 right zooms in the signal region, and it becomes clear that there is a tail of the double misID that may enter in the $K^0_S \to \mu^+\mu^-$ signal region.

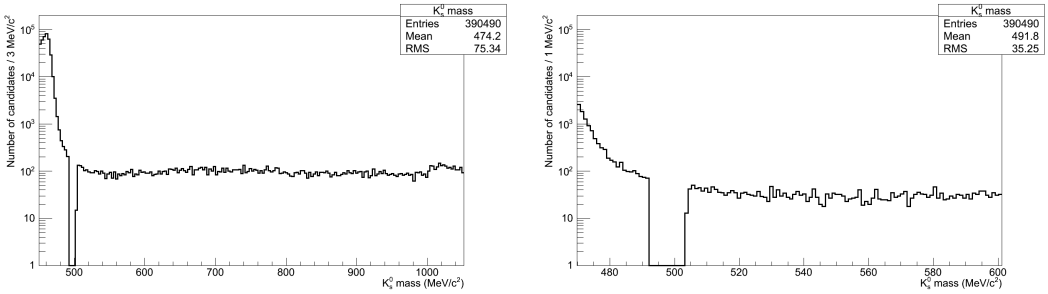

Figure 6.22: Mass distribution of $K^0_S \to \mu^+\mu^-$ after stripping and fiducial cuts. Left, whole mass range. Right, zoom in the signal region.

It should be also noticed that, even if the BDT is not effective for rejecting the double misID, it has an effect in the size of the tail entering in the signal region. Since the BDT is correlated with the mass resolution and also with the "quality" of the misID, as shown in figure 6.23, the mass resolution is better for high BDT bins, that meaning that the tail will also be less relevant. For the quality of the misID, high BDT double misID is composed mainly by candidates of very good quality, i.e., those in which the tracks are hardly affected by the decay in flight. This implies that the momentum resolution is also better, so the mass spectra less smeared (and therefore with lower probability to enter in the signal region).

Even when $K^0_S \to \pi^+\pi^-$ double misID has been shown to be non peaking in the signal region, it still has a tail that has to be studied in order to properly estimate the expected background. The 70 M MC10 sample has been used in order to find the right PDF to parameterise this tail. Due to the lack of statistics, the main study has been performed with the whole sample of $K^0_S \to \pi^+\pi^-$ with single misID, i.e, $K^0_S \to \pi^+\pi^-$ in which at least one of the pions suffers a decay in flight. Several different PDFs have been tried for the fit:

- Inv. polynomial (also called "power law" function, following the formula $\rho \sim 1/(m-m_0)^n$, where the free parameters are m_0 and n)
- Landau
- Breit-Wigner
- Simple exponential

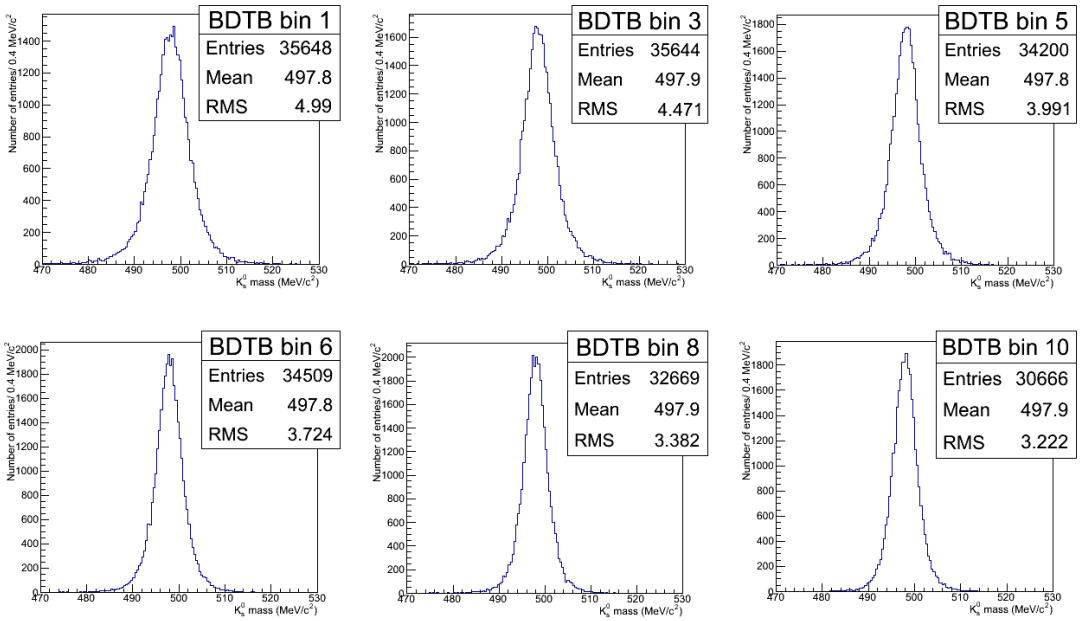

Figure 6.23: Mass resolution in different BDT bins. The larger the BDT, the better the resolution. The plot is built with $K^0_S \to \pi^+\pi^-$ from MB data. The BDT chosen is the one used in sample B TIS.

In a similar way, different starting points for the fit in the left hand sideband have been evaluated. The starting points of the fit have been chosen to be 465, 470 and 475 MeV/c^2. The fit in all the cases ends up in 550 MeV/c^2, skipping the 485-510 MeV/c^2 range, in which the background content is estimated.

Then, in order to choose the best among the different fits, the expected amount of candidates obtained with the fit has been compared with the actual amount of candidates in the signal region. Furthermore, the $\chi^2/ndof$ of all the fits has been obtained and compared as well. The $\chi^2/ndof$ has been evaluated in the background region, where the fit is actually done, and in the signal region, where the candidates which are seeked to estimate lie. Figure 6.24 shows the fits with all the explained variations. Table 6.10 shows the background expected by all the models, sorted according to the distance to the real value. Table 6.11 shows the $\chi^2/ndof$ of all the fits, sorted as well from the largest to the smallest probability in the signal region. The best predictions are given by the Breit-Wigner and Inv. polynomial starting at 470 MeV/c^2. As for the $\chi^2/ndof$, both models have large probabilities in the background region, but the Breit-Wigner has a poorer probability in the signal region.

The results in tables 6.10 and 6.11, together with some more systematic studies done in [202] lead to the choice of the power law PDF starting at 470 MeV/c^2 as the fitting model.

In summary, the double misID in $K^0_S \to \pi^+\pi^-$ is not peaking in the signal region of $K^0_S \to \mu^+\mu^-$, but it enters as a tail that can be parameterised with a power law function starting at 470 MeV/c^2. The specific parameters of the PDF will be fitted in data, in the different BDT bins, similarly to the $B^0_{(s)} \to \mu^+\mu^-$ case (section 5.8.4). Finally, even if in this study the signal region started at 485 MeV/c^2, the actual signal region in data will begin at 492 MeV/c^2, which also defines the starting point of the blind region. The reason to do so is trying to increase the area of the fit, since the 470-485 range was seen in data to be very poor in statistics in some of the BDT bins. For the same reason, the upper sideband will start at 504 MeV/c^2, instead of the 510 MeV/c^2 used in this MC study.

Model	M1	Prediction
Breit-Wigner	470	$58.26^{+5.70}_{-5.14}$
Inv. polynomial	470	$47.08^{+6.83}_{-5.88}$
Landau	475	$44.91^{+6.68}_{-5.91}$
Inv. polynomial	475	$43.35^{+7.39}_{-6.47}$
Breit-Wigner	475	$38.68^{+7.68}_{-6.76}$
Landau	470	$77.24^{+5.28}_{-4.95}$
Inv. polynomial	465	$28.36^{+4.39}_{-3.89}$
Exponential	475	$86.15^{+11.98}_{-11.34}$
Exponential	470	$19.10^{+3.44}_{-2.95}$
Exponential	465	$4.53^{+0.47}_{-0.43}$
Breit-Wigner	465	$152.73^{+5.18}_{-4.96}$
Landau	465	$289.34^{+7.24}_{-7.06}$

Table 6.10: Prediction of background in the signal region depending on different models and values of M1. M1 refers to the beginning of the left sideband used for fitting. The signal region spans the 485-510 MeV/c^2 range. The fitted region, the M1-485, 510-550 MeV/c^2 range. The results are sorted according to the distance to the actual measured value, $53.00^{+7.28}_{-7.28}$.

		Signal region			Fitting region		
Model	M1	χ^2	ndof	Prob.	χ^2	ndof	Prob.
Breit-Wigner	475	19.94	37.00	0.99	12.21	73.00	1.00
Inv. polynomial	475	21.31	37.00	0.98	12.20	73.00	1.00
Landau	475	21.59	37.00	0.98	12.64	73.00	1.00
Exponential	470	21.10	34.00	0.96	65.01	73.00	0.74
Inv. polynomial	465	20.61	31.00	0.92	71.29	73.00	0.53
Inv. polynomial	470	23.02	34.00	0.92	14.01	72.00	1.00
Breit-Wigner	470	30.59	34.00	0.64	18.74	72.00	1.00
Exponential	465	31.11	31.00	0.46	60.45	74.00	0.87
Landau	470	37.09	34.00	0.33	52.89	72.00	0.96
Exponential	475	50.27	37.00	0.07	32.46	74.00	1.00
Breit-Wigner	465	273.57	31.00	0.00	227.12	73.00	0.00
Landau	465	657.69	31.00	0.00	1009.27	73.00	0.00

Table 6.11: Quality of the fit in the signal and fitted regions depending on different models and values of M1. M1 refers to the beginning of the left sideband used for fitting. The signal region spans the 485-510 MeV/c^2 range. The fitted region, the M1-485, 510-550 MeV/c^2 range. The results are sorted according to the χ^2 probability in the signal region.

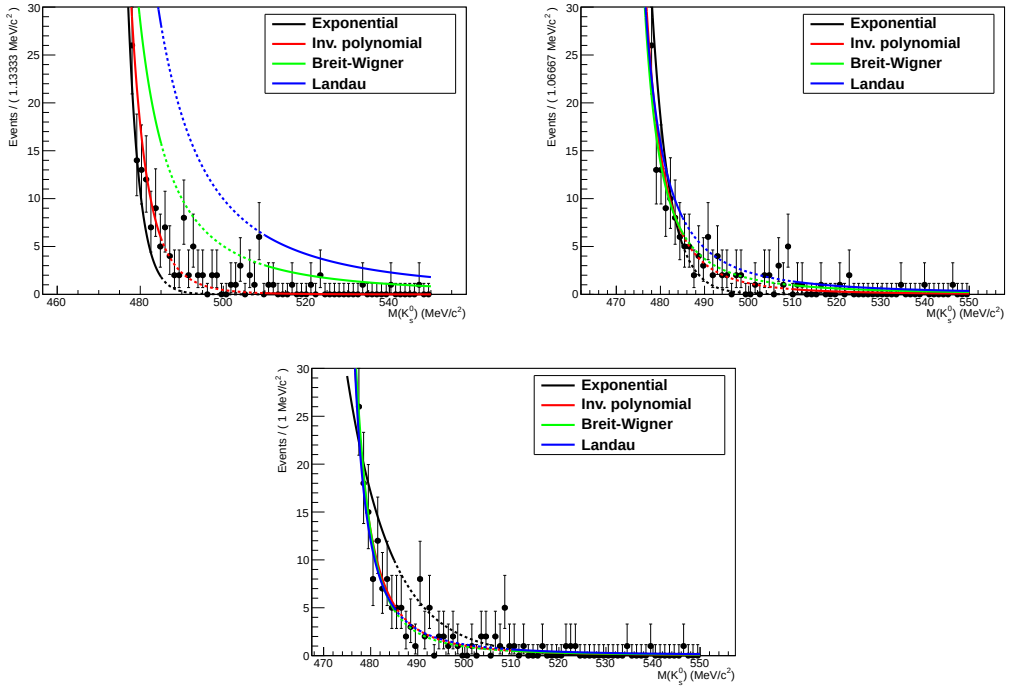

Figure 6.24: Fit results on the main $K^0_S \to \pi^+\pi^-$ MC10 misID sample. Each of the plots corresponds to a different starting point for the fitting region (M1, i.e., beginning of the left sideband). Top left, M1=465 MeV/c^2. Top right, M1=470 MeV/c^2. Bottom, M1=475 MeV/c^2. Black point represents the MC data, while different lines represent different fitting models. The fitted region spans the M1-485, 510-550 MeV/c^2 range, while a prediction on the amount of background is performed in the signal region, in the 485-510 MeV/c^2 range. The level of accuracy of the fits predicting the observed background is shown in table 6.10. The quality of the fits in both fitted and signal region is showed in table 6.11.

6.6.3 $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$

$K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$, with a $\mathcal{B}$ of $(4.69 \pm 0.05) \times 10^{-4}$ [5] can be a potentially dangerous background for $K^0_S \to \mu^+\mu^-$. Given the fact that the neutrino is never reconstructed, the initial K^0_S is built only with the pion and the muon. Therefore, the region of the phase space in which the momentum of the neutrino is smaller is the one with a largest probability to look like the signal. Within this region of the phase space, the cases in which the pion is misidentified as a muon may become close to $K^0_S \to \mu^+\mu^-$.

For making an initial study of this background, a large generator-level[5] MC sample has been used. This accounts for 1 M events of the type 34512100, presented in table 6.1, and generated with Gauss v41r2. For better understanding this background, it must be said that the K^0_S looked at for the analysis are those belonging to the prompt subsample, which is the more likely to survive the tight cut on the K^0_S IP (0.4 mm) imposed at stripping level, as seen in section 6.3. In this way, the fact that the K^0_S IP is reconstructed using only the pion and the muon results in very important differences between $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ and $K^0_S \to \mu^+\mu^-$. Similarly, this also happens for the K^0_S mass. Therefore, the initial study performed is focused in these two variables. The *fake* IP is obtained at MCTruth level, using only the pion and muon information. The mass, obtained as well using the MCTruth pion and muon, is also smeared with a Gaussian PDF using a resolution at the level of that of K^0_S measured with $K^0_S \to \pi^+\pi^-$, $\sigma = 3.7$ MeV/c^2 (see for instance figure 6.23).

[5] so without simulating neither the rest of the underlying event nor the interactions with the material.

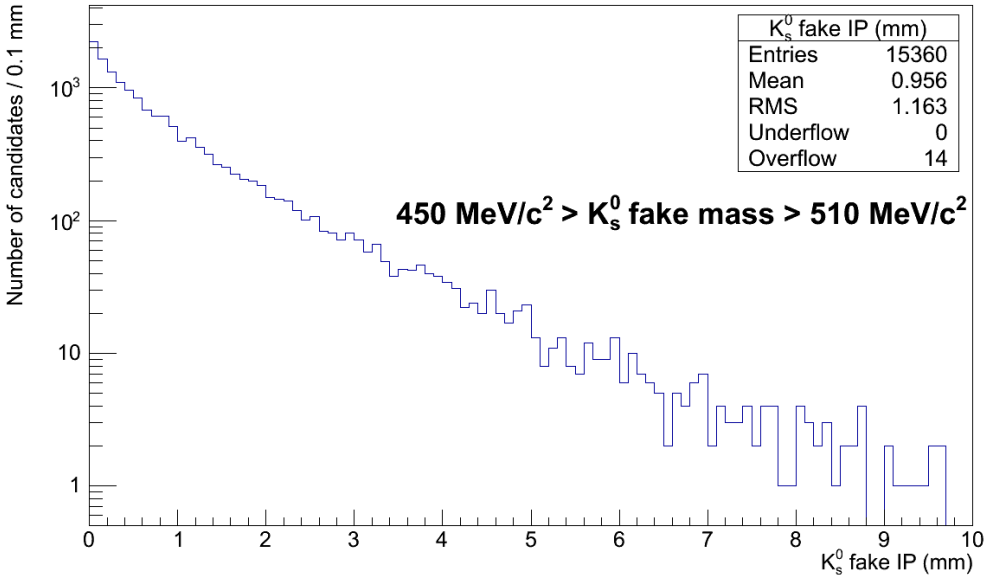

Figure 6.25: Fake IP distribution of the prompt K_S^0 calculated only with the pion and muon. The distribution is shown only for $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu$ in which the smeared fake mass is in the 450-510 MeV/c^2 range.

The fake IP distribution of the K_S^0 in $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu$, calculated only with the pion and the muon, can be seen in figure 6.25. In the figure, only the K_S^0 with a fake mass (obtained again only with the pion and muon, smeared to match the mass K_S^0 resolution and in the $\mu^+\mu^-$ hypothesis) in the range 450-510 MeV/c^2 range are shown. This range includes both the left sideband (450-492 MeV/c^2) of $K_S^0 \to \mu^+\mu^-$ and the signal region (492-504 MeV/c^2). The cut at stripping level is set at 0.4 mm, i.e., to the first bins of the plot. The cases in which the K_S^0 is reconstructed with low IP correspond to decays in which the neutrino momentum is very low, so that the pion and the muon are enough to accurately reproduce the momentum of the mother. For the same reason, in these cases the reconstructed mass becomes closer to the K_S^0 mass. This can be seen in figure 6.26, where the fake mass distribution is shown in different bins of fake K_S^0 IP (the fake mass and IP being obtained as explained before). As expected, the bin of lowest IP has the mass distribution closest to the one of K_S^0. The mass distribution in these plots is shown in the $\pi^+\mu^-$ and $\mu^+\mu^-$ hypothesis. It can be seen that changing from the first to the second hypothesis increases the separation between $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu$ and $K_S^0 \to \mu^+\mu^-$.

In order to estimate the expected yield of $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu$ in the ~ 1 fb^{-1} being analysed, a normalisation is done with respect to $K_S^0 \to \pi^+\pi^-$ TIS. This normalisation uses only MC information, and is done independently to obtain the expected $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu|_{TIS}$ and $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu|_{TOS}$. For this, full MC11 $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu$ and $K_S^0 \to \pi^+\pi^-$ samples are used. In the case of $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu$, two different samples were included, both presented in table 6.1. First, with dec type 34512100, corresponds to the the decay with no generator level cut. Second, with dec type 34512103, forces the decay in flight of the pion to a muon and a neutrino in a position so that $z < 11$ m. For $K_S^0 \to \pi^+\pi^-$, the sample is obtained stripping the minbias MC11 presented in table 6.2.

The formula used to determine the expected yield for $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu$ TIS (TOS) in 2011 data sample can be found in equation 6.6 (6.7). There, $\mathcal{B}(K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu)$ and $\mathcal{B}(K_S^0 \to \pi^+\pi^-)$ are the branching ratios of $K_S^0 \to \pi^+\pi^-$ and $K_S^0 \to \pi^+\mu^-\bar{\nu}_\mu$, ϵ^{SEL} the total reconstruction and selection efficiencies, $\epsilon^{TIS/SEL}$ and $\epsilon^{TOS/SEL}$ the TIS and TOS efficiencies in selected candidates and $N^{TIS}_{K_S^0\to\pi^+\pi^-;SelTIS}$ ($N^{TIS}_{K_S^0\to\pi^+\pi^-;SelTOS}$) is the total number of selected $K_S^0 \to \pi^+\pi^-$ TIS

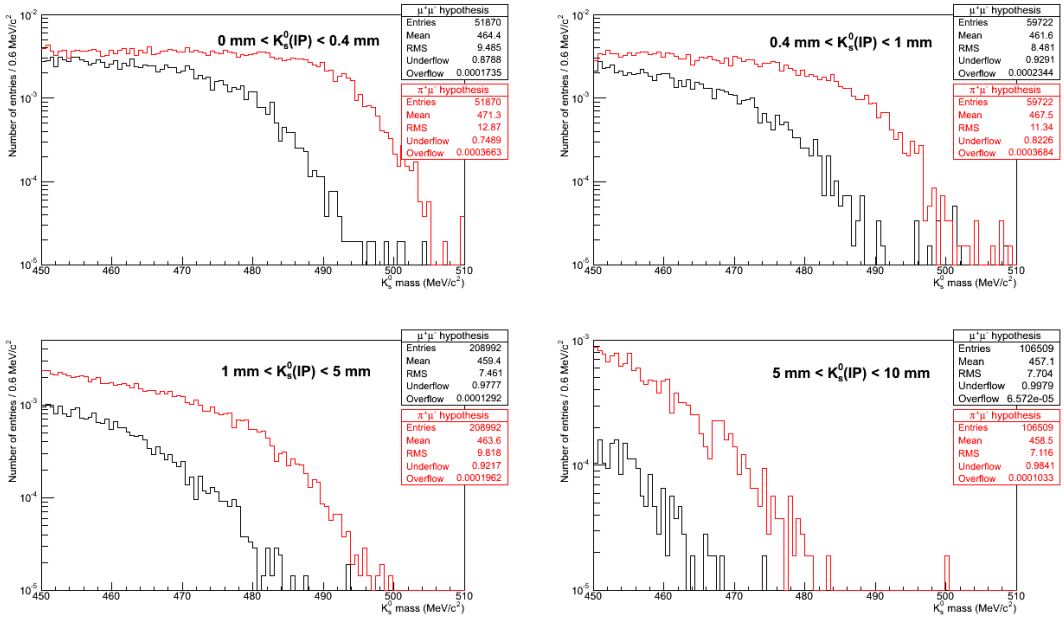

Figure 6.26: Fake mass distribution of the prompt K^0_S calculated only with the pion and muon momenta in different bins of the fake K^0_S IP. The K^0_S mass is shown in the $\pi^+\pi^-$ hypothesis (red curve) and $\mu^+\mu^-$ hypothesis (black curve). The mass is smeared following a Gaussian of $\sigma = 3.7$ MeV/c^2. The fake K^0_S IP is calculated with only the pion and muon.

expected in the 1 fb^{-1} sample, taking into account the prescale in the $K^0_S \to \pi^+\pi^-$ stripping line and using the $K^0_S \to \mu^+\mu^-$ cuts for the TIS (TOS) category, without the PID. The subscripts $\pi\pi$ and $\pi\mu\nu$ note the $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ decays. The selection includes the set of stripping cuts in table 6.3, the fiducial cuts explained in sections 6.3 and 6.4 (note that the fiducial cuts are different for the TIS and TOS category, this is indicated by the different superscripts $\epsilon^{SEL,TIS}$ and $\epsilon^{SEL,TOS}$) and the BDT larger than 0.2. Given the scope of this study, the BDT has been generically chosen as that of the TIS B sample. For $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$, muon PID cuts have been applied to both the pion and the muon. The mass range selected has been that of $K^0_S \to \mu^+\mu^-$ (which in this case is equivalent to $M(K^0_S) > 450$, as seen in figure 6.26).

$$N^{TIS}_{K^0_S\to\pi^+\mu^-\bar{\nu}_\mu} = \frac{\mathcal{B}(K^0_S \to \pi^+\mu^-\bar{\nu}_\mu)}{\mathcal{B}(K^0_S \to \pi^+\pi^-)} \times \frac{\epsilon^{SEL,TIS}_{\pi\mu\nu}}{\epsilon^{SEL,TIS}_{\pi\pi}} \times \frac{\epsilon^{TIS/SEL}_{\pi\mu\nu}}{\epsilon^{TIS/SEL}_{\pi\pi}} \times N^{TIS}_{K^0_S\to\pi^+\pi^-;SelTIS} \tag{6.6}$$

$$N^{TOS}_{K^0_S\to\pi^+\mu^-\bar{\nu}_\mu} = \frac{\mathcal{B}(K^0_S \to \pi^+\mu^-\bar{\nu}_\mu)}{\mathcal{B}(K^0_S \to \pi^+\pi^-)} \times \frac{\epsilon^{SEL,TOS}_{\pi\mu\nu}}{\epsilon^{SEL,TOS}_{\pi\pi}} \times \frac{\epsilon^{TOS/SEL}_{\pi\mu\nu}}{\epsilon^{TIS/SEL}_{\pi\pi}} \times N^{TIS}_{K^0_S\to\pi^+\pi^-;SelTOS} \tag{6.7}$$

As explained before, and given the very low selection efficiencies for $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ with the mentioned cuts, a MC with a generation cut forcing the decay in flight of the pion has been used to increase the amount of offline selected candidates. This MC has the dec type 34512103. However, in order to obtain the correct ϵ^{SEL}, the real MCTruth probability for the pion decay in flight has to be used. This is obtained from the sample with no generation cuts, with dec type 34512100. This is all shown in equation 6.8 for the TIS category. The total number of offline selected candidates with the TIS cuts in sample 34512103 is 2. For the TOS category, a very tight cut is also imposed in the maximum p_T of both daughters ($max(p_{T1}, p_{T2}) > 1300$).

Since this cut gives no offline candidates in $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ 34512103 sample, the efficiency of this cut has been assessed at the production level for both $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ (in the $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ case after applying the cuts in the fake IP and mass seen before). With that, the ratio $\epsilon^{SEL,TOS}$ has been obtained as a correction to the ratio of $\epsilon^{SEL,TIS}$, as seen in equation 6.9.

$$\epsilon^{SEL,TIS}_{\pi\mu\nu} = \frac{N^{SEL,TIS}_{\pi\mu\nu}}{N^0_{\pi\mu\nu}} = \left.\frac{N^{SEL,TIS}_{\pi\mu\nu}}{N^{\pi\ dcf}_{\pi\mu\nu}}\right|_{34512103} \times \left.\frac{N^{\pi\ dcf}_{\pi\mu\nu}}{N^0_{\pi\mu\nu}}\right|_{34512100} \tag{6.8}$$

$$\frac{\epsilon^{SEL,TOS}_{\pi\mu\nu}}{\epsilon^{SEL,TOS}_{\pi\pi}} = \frac{\epsilon^{SEL,TIS}_{\pi\mu\nu}}{\epsilon^{SEL,TIS}_{\pi\pi}} \times \frac{\epsilon^{MAX\ p_T}_{\pi\mu\nu;PROD[IP(K^0_S)<0.4;M(K^0_S)>450]}}{\epsilon^{MAX\ p_T}_{\pi\pi;PROD}} \tag{6.9}$$

Finally, for the ratio of trigger efficiencies, two different approaches have been considered. For the ratio of TIS efficiencies, it has simply been considered to be 1. For the ratio of TOS/TIS efficiencies, the formula in equation 6.10 has been used. This formula uses the ratio of TOS and TIS efficiencies from MC11 $K^0_S \to \mu^+\mu^-$ (noted as $\mu\mu$ in the equation) and corrects it by the different TOS efficiencies of the misidentified pions and real muons. For obtaining this last correction, the same MC10 from section 6.4.2 is used. This included a large set of misidentified pions from $K^0_S \to \pi^+\pi^-$.

$$\frac{\epsilon^{TOS/SEL}_{\pi\mu\nu}}{\epsilon^{TIS/SEL}_{\pi\pi}} = \frac{\epsilon^{TOS/SEL}_{\pi\mu\nu}}{\epsilon^{TIS/SEL}_{\mu\mu}} = \frac{\epsilon^{TOS/SEL}_{\mu\mu}}{\epsilon^{TIS/SEL}_{\mu\mu}} \times \frac{\epsilon^{TOS/SEL}_{\pi\ misID}}{\epsilon^{TOS/SEL}_{\mu}} \tag{6.10}$$

A summary of the numbers entering in the calculation of the yield of $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ TIS and TOS can be found in table 6.12. The two dominant errors for these calculations come from the ratios $\frac{\epsilon^{TOS/SEL}_{\pi\mu\nu}}{\epsilon^{TIS/SEL}_{\pi\pi}}$ and $\frac{\epsilon^{SEL,TIS}_{\pi\mu\nu}}{\epsilon^{SEL,TIS}_{\pi\pi}}$. The rest of the errors have been neglected. With all this, the expected number of $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ TIS (TOS) in the range $M(K^0_S) > 450$ is 6^{+7}_{-4} (7^{+20}_{-5}).

It must be said that the numbers just quoted include the whole mass window of $K^0_S \to \mu^+\mu^-$, and are expected to lie mainly in the left side of it, as it was seen in figure 6.26. Some shift towards the $K^0_S \to \mu^+\mu^-$ signal region could arise, given the fact that the BDT cut selects those $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ looking closer to a two body decay K^0_S, but the available MC statistics do not allow further studies (the two mentioned candidates surviving the offline cuts have masses below 470 MeV/c^2). The surviving events are also expected to have low values of BDT, and must be splitted in the 20 BDT bins for each category (TIS/TOS). In summary, this background is considered to be negligible in practice when compared to the $K^0_S \to \pi^+\pi^-$ double misID and combinatorial background, although it may become more relevant for determining a BR of $K^0_S \to \mu^+\mu^-$ close to the SM prediction.

As a final remark, a very similar channel that has also been looked at as a potential background in the search for $K^0_S \to \mu^+\mu^-$ is $K^0_L \to \pi^+\mu^-\bar{\nu}_\mu$. Even if this decay is expected to occur mainly outside VELO (because of the large K^0_L lifetime), the fact that $\mathcal{B}(K^0_L \to \pi^+\mu^-\bar{\nu}_\mu) = 0.27$ [5] is much larger than that of $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ makes necessary a more careful calculation. Assuming the same ratio of reconstruction efficiencies for $K^0_L \to \pi^+\mu^-\bar{\nu}_\mu$ and $K^0_L \to \mu^+\mu^-$ with respect to the K^0_S modes with the same final states, $\sim 2.3 \times 10^{-3}$ as obtained in section 6.6.1, and using also $\mathcal{B}(K^0_S \to \pi^+\mu^-\bar{\nu}_\mu) = 4.6 \times 10^{-4}$, the effective $\mathcal{B}(K^0_L \to \pi^+\mu^-\bar{\nu}_\mu)$ can be seen to be:

$$\mathcal{B}(K^0_L \to \pi^+\mu^-\bar{\nu}_\mu)|_{eff} \sim 1.35 \times \mathcal{B}(K^0_S \to \pi^+\mu^-\bar{\nu}_\mu)|_{eff}$$

With this, $K^0_L \to \pi^+\mu^-\bar{\nu}_\mu$ would be at the level of $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$, so also negligible for the current dataset. Once again, this background would need to be revisited for a search for $K^0_S \to \mu^+\mu^-$ at the level of the SM prediction.

$\mathcal{B}(K^0_S \to \pi^+\mu^-\bar{\nu}_\mu)$	4.69×10^{-4}
$\mathcal{B}(K^0_S \to \pi^+\pi^-)$	69.2×10^{-2}
$\frac{N^{SEL,TIS}_{\pi\mu\nu}}{N^{\pi\ dcf}_{\pi\mu\nu}}\Big\vert_{34512103}$	0.0961
$\frac{\epsilon^{SEL,TIS}_{\pi\mu\nu}}{\epsilon^{SEL,TIS}_{\pi\pi}}$	$8.13^{+17.29}_{-3.01} \times 10^{-5}$
$\frac{\epsilon^{TIS/SEL}_{\pi\mu\nu}}{\epsilon^{TIS/SEL}_{\pi\pi}}$	1
$N^{TIS}_{K^0_S \to \pi^+\pi^-;SelTIS}$	113522000
$N^{TIS}_{K^0_S \to \pi^+\mu^-\bar{\nu}_\mu}$	6^{+7}_{-4}
$\frac{\epsilon^{MAX\ p_T}_{\pi\mu\nu;PROD[IP(K^0_S)>0.4;M(K^0_S)>450]}}{\epsilon^{MAX\ p_T}_{\pi\pi;PROD}}$	0.4563
$\frac{\epsilon^{SEL,TOS}_{\pi\mu\nu}}{\epsilon^{SEL,TOS}_{\pi\pi}}$	$3.71^{+7.89}_{-1.37} \times 10^{-5}$
$\frac{\epsilon^{TOS/SEL}_{\pi\ misID}}{\epsilon^{TOS/SEL}_{\mu}}$	0.2275
$\frac{\epsilon^{TOS/SEL}_{\pi\mu\nu}}{\epsilon^{TIS/SEL}_{\pi\pi}}$	$62.80^{+110.22}_{-39.57}$
$N^{TIS}_{K^0_S \to \pi^+\pi^-;SelTOS}$	4612000
$N^{TOS}_{K^0_S \to \pi^+\mu^-\bar{\nu}_\mu}$	7^{+20}_{-5}

Table 6.12: Value of some of the parameters entering in the computation of $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ yield (see text for details)

6.6.4 $K^0_S \to \mu^+\mu^-\gamma$

Similarly to $B^0_s \to \mu^+\mu^-\gamma$ in the previous analysis, $K^0_S \to \mu^+\mu^-\gamma$ could become a relevant background for $K^0_S \to \mu^+\mu^-$ whenever the photon does not come from the FSR. Due to the momentum taken by the photon, the reconstructed $K^0_S \to \mu^+\mu^-\gamma$ candidates should mainly lie in the left mass sideband of $K^0_S \to \mu^+\mu^-$.

The prediction for $\mathcal{B}(K^0_S \to \mu^+\mu^-\gamma)/\mathcal{B}(K^0_S \to \gamma\gamma)$ is 2.8×10^{-4} [203]. Together with the measurement of the $\gamma\gamma$ mode ($[2.63 \pm 0.17] \times 10^{-6}$) [5], $\mathcal{B}(K^0_S \to \mu^+\mu^-\gamma)$ can be obtained to be $\sim 7.4 \times 10^{-10}$. As it will be seen, this $\mathcal{B}$ is below the sensitivity for $K^0_S \to \mu^+\mu^-$ in this analysis.

It should be also remarked that the BDT, trigger and mass window are expected to further suppress this background.

6.7 Normalisation

This section is devoted to the normalisation, which permits converting an observed number of $K^0_S \to \mu^+\mu^-$ candidates in a branching fraction. As already explained, the normalisation will be independent for TIS and TOS candidates.

The $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ is computed using:

$$\frac{\mathcal{B}(K^0_S \to \mu^+\mu^-)}{\mathcal{B}(K^0_S \to \pi^+\pi^-)} = \frac{\epsilon_{\pi\pi}}{\epsilon_{\mu\mu}} \frac{N_{K^0_S \to \mu^+\mu^-}}{N_{K^0_S \to \pi^+\pi^-}} \tag{6.11}$$

Where ϵ are the efficiencies in each BDT bin, which are separated as:

- $\epsilon^{SEL/REC}$, which is the offline efficiency, including reconstruction and selection but excluding, for convenience, the Particle Identification efficiency. $\epsilon^{SEL/REC}$ can be explained as the probability for a $K^0_S \to \pi^+\pi^-$ ($K^0_S \to \mu^+\mu^-$) decay generated in a pp collision, to have been reconstructed and selected in the case of a 100% efficient trigger: $\epsilon^{SEL/REC} = \frac{N^{SEL}}{N^{GEN4\pi}}$.

- ϵ^{PID}, which is the efficiency of the muon identification of $K^0_S \to \mu^+\mu^-$, $\epsilon^{PID} = \frac{N^{SEL\&PID}}{N^{SEL}}$.

- $\epsilon^{TRIG/SEL}$ which is the efficiency of the trigger for events that would be offline selected, $\epsilon^{TRIG/SEL} = \frac{N^{SEL\&PID\&TRIG}}{N^{SEL\&PID}}$. In a similar way, the TOS and TIS efficiencies can be defined as $\epsilon^{TOS/SEL} = \frac{N^{SEL\&PID\&TOS}}{N^{SEL\&PID}}$ and $\epsilon^{TIS/SEL} = \frac{N^{SEL\&PID\&TIS}}{N^{SEL\&PID}}$. Under this definition, the TISTOS formalism can be used to get trigger efficiencies from data, as described in [179] and used in section 5.9.5.

The muonID efficiency is evaluated in bins of BDT (both for the TOS and TIS case) using the momentum and transverse momentum spectra of the $K^0_S \to \pi^+\pi^-$ on MB data, which is trigger unbiased. For the ratio of trigger efficiencies, slightly different strategies are considered for the TOS and TIS samples. For the TOS case, the normalisation is done to $K^0_S \to \pi^+\pi^-$ after GECs in minimum bias events. The way to compute this number will be explained later. For the TIS case, the normalisation is done to $K^0_S \to \pi^+\pi^-$ TIS. In summary:

$$\frac{\mathcal{B}(K^0_S \to \mu^+\mu^-)}{\mathcal{B}(K^0_S \to \pi^+\pi^-)} = \frac{\epsilon^{SEL/REC}_{\pi\pi}}{\epsilon^{SEL/REC}_{\mu\mu}} \times \frac{1}{\epsilon^{PID}_{\mu\mu}} \times \frac{\epsilon^{TIS/SEL}_{\pi\pi}}{\epsilon^{TIS/SEL}_{\mu\mu}} \times \frac{N^{TIS}_{K^0_S \to \mu^+\mu^-}}{N^{TIS}_{K^0_S \to \pi^+\pi^-}} \tag{6.12}$$

$$\frac{\mathcal{B}(K^0_S \to \mu^+\mu^-)}{\mathcal{B}(K^0_S \to \pi^+\pi^-)} = \frac{\epsilon^{SEL/REC}_{\pi\pi}}{\epsilon^{SEL/REC}_{\mu\mu}} \times \frac{1}{\epsilon^{PID}_{\mu\mu}} \times \frac{s^{MB}}{\epsilon^{TOS/SEL}_{\mu\mu}} \times \frac{N^{TOS}_{K^0_S \to \mu^+\mu^-}}{N^{MB,GECS}_{K^0_S \to \pi^+\pi^-}} \tag{6.13}$$

Being s^{MB} a scaling factor from the MB stream to the physics trigger streams.

The quantities:

$$\alpha_{TIS} = \frac{\epsilon^{SEL/REC}_{\pi\pi}}{\epsilon^{SEL/REC}_{\mu\mu}} \times \frac{1}{\epsilon^{PID}_{\mu\mu}} \times \frac{\epsilon^{TIS/SEL}_{\pi\pi}}{\epsilon^{TIS/SEL}_{\mu\mu}} \times \frac{1}{N^{TIS}_{K^0_S\to\pi^+\pi^-}} \tag{6.14}$$

$$\alpha_{TOS} = \frac{\epsilon^{SEL/REC}_{\pi\pi}}{\epsilon^{SEL/REC}_{\mu\mu}} \times \frac{1}{\epsilon^{PID}_{\mu\mu}} \times \frac{s^{MB}}{\epsilon^{TOS/SEL}_{\mu\mu}} \times \frac{1}{N^{MB,GECS}_{K^0_S\to\pi^+\pi^-}} \tag{6.15}$$

are hereafter called *normalisation factors* and are calculated in bins of BDT. For a given number N of observed/excluded $K^0_S \to \mu^+\mu^-$ events, the corresponding $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ is then $\alpha \times N$.

6.7.1 Offline efficiencies

The ratio of reconstruction and selection efficiencies is evaluated using MC simulation. As the momentum spectrum of the K^0_S in data is different to that of the simulation [204], the ratio of efficiencies is calculated in bins of transverse momentum (p_T) and rapidity y[6] of the K^0_S, and the ratio of efficiencies is then reweighted according to the y, p_T distributions in the mentioned reference. However [204] only covers up to $p_T = 2.5\,\text{GeV}/c$ while in the current analysis (specially in the TOS case), there is a sizeable fraction of K^0_S with higher p_T. As the ratio of efficiencies varies very little with p_T (specially for high p_T), the effect of neglecting the cross section correction at high p_T is expected to be small. To cross check this, a naive differential cross section is estimated in data just by:

$$\sigma^{K^0_S}_{data,i} \sim \frac{N^{K^0_S\to\pi^+\pi^-}_{MBdata,i}}{\epsilon^{K^0_S\to\pi^+\pi^-}_{MC,i}} \tag{6.16}$$

which allows the extension of the p_T range up to higher values. A good agreement is found in general between the two approaches.

Table 6.13 shows the ratio of selection efficiencies, in bins of y, p_T.

Table 6.14 (table 6.15) shows this ratio in bins of BDT for the TIS (TOS) analysis in sample A. The robustness of the ratios against the different estimations of the differential cross sections are also shown. Table 6.16 shows the ratio of selection efficiencies in bins of BDT for the sample B. Before any selection cut, the reconstruction efficiency for $K^0_S \to \pi^+\pi^-$ is found to be 64% of that of $K^0_S \to \mu^+\mu^-$.

y	p_T (MeV/c)		
	150.0 – 650.0	650.0 – 1000.0	1000.0 – 2500.0
2.0 – 2.5	0.615 ± 0.069	0.647 ± 0.04	0.68 ± 0.036
2.5 – 3.0	0.619 ± 0.027	0.701 ± 0.023	0.708 ± 0.028
3.0 – 3.5	0.709 ± 0.024	0.815 ± 0.027	0.692 ± 0.031
3.5 – 4.0	0.722 ± 0.024	0.732 ± 0.023	0.774 ± 0.034
4.0 – 4.5	0.705 ± 0.046	0.672 ± 0.036	0.599 ± 0.04

Table 6.13: Selection efficiency ratio $\epsilon_{\pi\pi}/\epsilon_{\mu\mu}$ in bins of transverse momentum (p_T) and rapidity (y).

[6] $y = \frac{1}{2}\log(\frac{E+p_z}{E-p_z})$.

BDT bin	$\frac{\epsilon^{SEL}_{\pi\pi}}{\epsilon^{SEL}_{\mu\mu}}(MD)$	$\frac{\epsilon^{SEL}_{\pi\pi}}{\epsilon^{SEL}_{\mu\mu}}(MU)$	$\frac{\epsilon^{SEL}_{\pi\pi}}{\epsilon^{SEL}_{\mu\mu}}(ext)$
1	0.62 ± 0.02	0.62 ± 0.02	0.63 ± 0.02
2	0.70 ± 0.03	0.71 ± 0.03	0.71 ± 0.03
3	0.70 ± 0.04	0.70 ± 0.04	0.70 ± 0.04
4	0.75 ± 0.05	0.75 ± 0.05	0.75 ± 0.05
5	0.73 ± 0.06	0.73 ± 0.06	0.73 ± 0.06
6	0.63 ± 0.04	0.63 ± 0.04	0.63 ± 0.04
7	0.76 ± 0.07	0.76 ± 0.07	0.77 ± 0.08
8	0.83 ± 0.12	0.83 ± 0.12	0.82 ± 0.11
9	0.82 ± 0.12	0.82 ± 0.12	0.79 ± 0.11
10	0.77 ± 0.05	0.77 ± 0.05	0.77 ± 0.05

Table 6.14: Ratio of selection efficiencies in bins of BDT for the TIS analysis in sample A. Three estimations are provided, accordingly to the y, p_T distribution assumed for the K^0_S at the production level. MD (MU): Using the results of Magnet Down (Up) from [204]. *ext*: Using an approximate differential cross section obtained from MB data, taking absolute offline efficiency from MC.

BDT bin	$\frac{\epsilon^{SEL}_{\pi\pi}}{\epsilon^{SEL}_{\mu\mu}}(MD)$	$\frac{\epsilon^{SEL}_{\pi\pi}}{\epsilon^{SEL}_{\mu\mu}}(MU)$	$\frac{\epsilon^{SEL}_{\pi\pi}}{\epsilon^{SEL}_{\mu\mu}}(ext)$
1	0.95 ± 0.30	0.95 ± 0.30	0.93 ± 0.30
2	0.58 ± 0.18	0.58 ± 0.18	0.56 ± 0.17
3	0.80 ± 0.31	0.80 ± 0.31	0.78 ± 0.31
4	1.15 ± 0.41	1.15 ± 0.41	1.10 ± 0.39
5	0.51 ± 0.19	0.51 ± 0.19	0.54 ± 0.21
6	0.75 ± 0.28	0.75 ± 0.28	0.71 ± 0.26
7	1.54 ± 0.91	1.54 ± 0.91	1.52 ± 0.94
8	0.58 ± 0.21	0.58 ± 0.21	0.54 ± 0.19
9	0.60 ± 0.24	0.61 ± 0.24	0.65 ± 0.30
10	0.70 ± 0.11	0.70 ± 0.11	0.69 ± 0.11

Table 6.15: Ratio of selection efficiencies in bins of BDT for the TOS analysis in sample A. Three estimations are provided, accordingly to the y, p_T distribution assumed for the K^0_S at the production level. MD (MU): Using the results of Magnet Down (Up) from [204]. *ext*: Using an approximate differential cross section obtained from MB data, taking absolute offline efficiency from MC.

BDT bin	$\frac{\epsilon^{SEL}_{\pi\pi}}{\epsilon^{SEL}_{\mu\mu}}(TISB)$	$\frac{\epsilon^{SEL}_{\pi\pi}}{\epsilon^{SEL}_{\mu\mu}}(TOSB)$
1	0.61 ± 0.02	0.69 ± 0.13
2	0.70 ± 0.03	1.31 ± 0.52
3	0.65 ± 0.03	0.67 ± 0.19
4	0.66 ± 0.04	0.49 ± 0.17
5	0.78 ± 0.06	0.85 ± 0.35
6	0.72 ± 0.09	1.49 ± 1.00
7	0.77 ± 0.06	0.56 ± 0.16
8	0.72 ± 0.06	1.14 ± 0.47
9	0.72 ± 0.07	1.09 ± 0.85
10	0.75 ± 0.06	1.63 ± 0.67

Table 6.16: Ratio of selection efficiencies in bins of BDT for the TIS and for the TOS analyses in sample B.

6.7.2 MuonID efficiency

As explained, the muonID efficiency appears in the normalisation factor after reconstruction and selection efficiencies. In order to obtain the correct sample to measure this efficiency at the mentioned stage, the $K^0_S \to \pi^+\pi^-$ MB data sample is used, as it is biased by all the mentioned factors but not by the trigger. The muonID refers to the product of the muon system acceptance and the muon identification efficiency (IsMuon requirement and Combined DLL larger than -4).

The procedure to obtain the muonID efficiency is convoluting a set of $\epsilon_{acceptance} \times \epsilon_{IsMuon} \times \epsilon_{CDLL>-4}$ curves as a function of p/p_T with the p/p_T spectra of $K^0_S \to \pi^+\pi^-$. This therefore assumes roughly the same spectra for $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu^+\mu^-$.

The $\epsilon_{acceptance} \times \epsilon_{IsMuon}$ curves as a function of p/p_T are the same used for $B^0_{(s)} \to \mu^+\mu^-$ analysis, obtained in section 5.5. As a reminder, figure 5.10 showed the efficiency curves and numerical values as a function of p, p_T.

For the $\epsilon_{CDLL>-4}$, the plot in figure 6.27 uses the same procedure and sample than shown before for the $\epsilon_{acceptance} \times \epsilon_{IsMuon}$. This plots shows the $\epsilon_{acceptance} \times \epsilon_{IsMuon}$ and $\epsilon_{acceptance} \times \epsilon_{IsMuon} \times \epsilon_{CDLL>-4}$ only in bins in p (integrating over p_T), indicating that the efficiency of the $CDLL$ cut with respect to IsMuon is almost 100%.

In order to map the $\epsilon_{CDLL>-4}$ as a function of p/p_T, a $B^+ \to J/\psi K^+$ trigger unbiased[7] data sample has been used. The sample has been selected with the same cuts seen in section 5.4. The map, calculated with respect to muons fulfilling IsMuon, has been obtained using the same bins than in figure 5.10. In order to obtain the curves, two fits to the $J/\psi \to \mu^+\mu^-$ mass spectra have been performed for each bin. One for the positive muons with $CDLL > -4$ and other for the positive muons with $CDLL < -4$. With the obtained number of J/ψ, the efficiency is obtained with $\epsilon_{CDLL>-4} = \frac{N(J/\psi)_{CDLL>-4}}{N(J/\psi)_{CDLL>-4}+N(J/\psi)_{CDLL<-4}}$. This method has the advantage of allowing the error propagation to correctly take into account the correlations. In order to obtain a systematic error, two fitting models are used. First, more complex, includes a third order Chebyshev polynomial for the background and a Crystal Ball for the signal. The n and a parameters of the Crystal Ball were taken as a function of σ using the parameterisation in [205]. Second fit uses a background only exponential model, obtaining the number of J/ψ from the difference in the total number of candidates and the background candidates given by the fit. The signal mass regions in both cases are slightly different (3050-3150 MeV/c^2 for the first and 3010-3160 MeV/c^2 for the second). The central value quoted uses the second fit, while the systematic error is the difference between the efficiencies obtained with both models. A example of the fits used can be seen in figure 6.28. The fact that the cut is so efficient with respect to IsMuon (see figure 6.27) makes particularly difficult the computation of the number of J/ψ with $CDLL < -4$ in some of the bins. Whenever the total number of candidates is smaller than 100, the number of candidates in the corresponding signal region is taken as the number of J/ψ. The error arising from this approximation is neglected for the analysis, since it gives anyway yields of $\sim 100\%$ and it affects mainly bins not important for the $K^0_S \to \mu^+\mu^-$ case. A list of the $\epsilon_{CDLL>-4}$ values obtained can be found in table 6.17.

	$(0.05 < p_T < 1.7)$	$(1.7 < p_T < 3.0)$	$(3.0 < p_T < 5.0)$	$(5.0 < p_T < 10.0)$
$(3 < p < 10)$	(99.49 ± 0.16) %	(99.60 ± 0.09) %	(100.00 ± 1.50) %	-
$(10 < p < 21)$	(99.62 ± 0.05) %	(99.92 ± 0.02) %	(99.95 ± 0.04) %	(100.00 ± 0.40) %
$(21 < p < 30)$	(99.81 ± 0.03) %	(99.93 ± 0.03) %	(99.97 ± 0.03) %	(100.00 ± 0.06) %
$(30 < p < 40)$	(99.91 ± 0.04) %	(99.92 ± 0.03) %	(100.00 ± 0.02) %	(100.00 ± 0.04) %
$(40 < p < 60)$	(99.95 ± 0.03) %	(99.94 ± 0.03) %	(99.97 ± 0.02) %	(100.00 ± 0.02) %

Table 6.17: $\epsilon_{CDLL>-4}$ vs. p/p_T with respect to IsMuon obtained from a set of trigger unbiased muons from $B^+ \to J/\psi K^+$. All the bins are defined in GeV/c. The errors quoted include a statistical and systematical contribution (see text for details).

In order to obtain the correct p/p_T spectra, the fiducial cuts (except, obviously, for the

[7] TIS with respect to the three trigger levels.

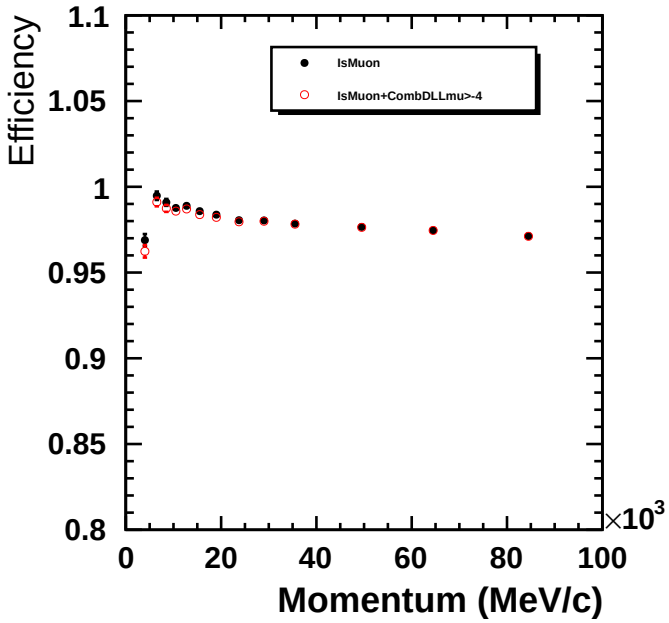

Figure 6.27: $\epsilon_{acceptance} \times \epsilon_{IsMuon}$ and $\epsilon_{acceptance} \times \epsilon_{IsMuon} \times \epsilon_{CDLL>-4}$ vs. p obtained from a set of unbiased muons from detached J/ψ.

muonID cuts), on top of the stripping (see section 6.3), are applied to the $K^0_S \to \pi^+\pi^-$ MB data sample. As the normalisation is done in bins of BDT, the muonID efficiency is also obtained for the corresponding bins, both in BDTA (in sample A) and in BDTB (in sample B) for the TIS and TOS categories. Figure 6.29 shows, as an example, the p and p_T distributions for three different TIS BDT bins for the MC11 $K^0_S \to \mu^+\mu^-$ and $K^0_S \to \pi^+\pi^-$. The distributions are very similar in both channels, which validates the use of $K^0_S \to \pi^+\pi^-$ MB for obtaining the efficiencies for $K^0_S \to \mu^+\mu^-$. For the TOS category, the computation of this table is simply informative, since the total TOS and muonID efficiency will be computed together, as it will be seen later. To remove any remanent background, the K^0_S are restricted to the signal region (492-504 MeV/c^2). The total muonID efficiency per bin is obtained combining the information in figure 5.10 and table 6.17.

After all the mentioned cuts, the $K^0_S \to \pi^+\pi^-$ sample is left with almost no background (see for instance figure 6.33). So the total efficiency is simply measured, for every BDT bin, by calculating the efficiency for each K^0_S candidate (using the $\pi^\pm$ p and p_T and the values in figure 5.10 and table 6.17) and then obtaining the mean (see equation 6.17).

$$\epsilon^{muonID}_{BDT\ bin\ i} = \left(\sum_{j}^{N^{cands}_{BDT\ bin\ i}} \epsilon_{muonID}(p_{j1}, p_{Tj1}) \times \epsilon_{muonID}(p_{j2}, p_{Tj2}) \right) / N^{cands}_{BDT\ bin\ i} \qquad (6.17)$$

The total lower and upper errors are calculated in a similar way, changing for every p/p_T bin the initial value by the value $\pm$ the error. The errors obtained with this procedure turn out to be very symmetrical. When a $\pi^\pm$ from the K^0_S has a p or p_T out of the defined binning, the closest bin is used. All the tracks with $p < 3\,\text{GeV}/c^2$ are considered with 0 efficiency, since this is the minimum momentum required by the IsMuon condition.

In order to obtain a complete systematic error for this result, the procedure is repeated in MC11 $K^0_S \to \mu^+\mu^-$ for each BDT bin in the TIS category. For the TOS category, given the low statistics after the fiducial cut in $max(p_{T1}, p_{T2}) > 1300$ MeV/c, the procedure is repeated after the BDT cut of the first BDT bin, and the systematic obtained assigned to all the BDT bins. With the known difference in the p_T spectra of K^0_S in data and MC, the results after applying the

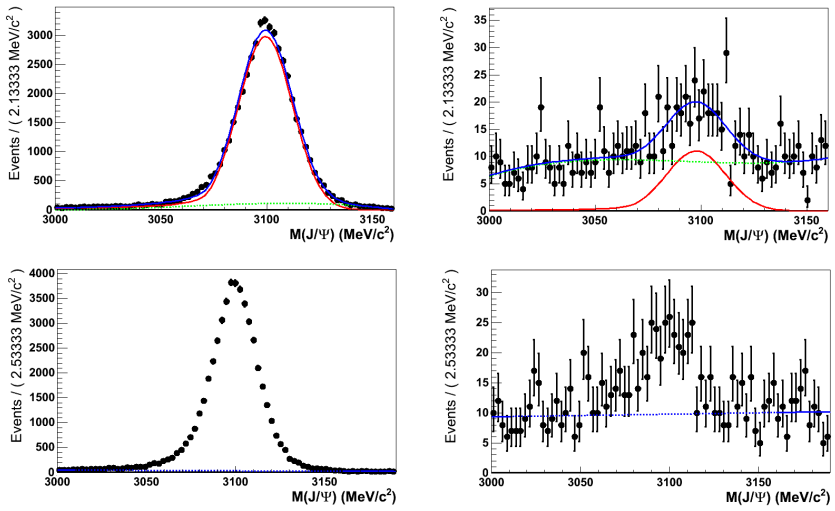

Figure 6.28: Fits for $\epsilon_{CDLL>-4}$ computation. This example corresponds to the p, p_T range of $3-10$ GeV/c and $0.05-1.7$ GeV/c. Top plots correspond to the complex fit model (third order Chebyshev polynomial for the background and a Crystal Ball for the signal), while bottom to the simple fit model (background only exponential). Left plots are obtained after applying $CDLL > -4$ to the μ^+, while right after applying $CDLL < -4$ to it.

weighting cannot be directly compared to the equivalent in data. However, the result obtained in MC can be related to the one obtained using the MC Truth information (simply by counting number of events before and after the cuts). Since the muonID efficiency is known to be very similar in data and MC (see for instance figures 5.8 and 5.9), this is a sensible way of measuring any inaccuracy due to the method just reported. The difference between both results in MC is then assigned as a systematic error in data. This error will become dominant with respect to the one arising from the errors in ϵ_{muonID} tables, and it will be the only one considered. The method just reported is almost equivalent to the one used in the $B^0_{(s)} \to \mu^+\mu^-$ analysis.

The results obtained with the explained method are shown in table 6.18 (6.19) for the different BDTA and BDTB bins in the TIS (TOS) category. The result is quite consistent throughout all the bins. The fact that the efficiencies are larger in the TOS category arises from the fact that the p, p_T spectra becomes harder there due to the fiducial cut in $max(p_{T1}, p_{T2}) > 1300$ MeV/c.

6.7.3 $K^0_S \to \mu^+\mu^-$ TOS efficiency

For the TOS category, given the fact that only single muon TOS candidates are considered, a similar mapping to the one seen for the muonID case can be applied. Furthermore, taking into account the fiducial cuts in $max(p_{T1}, p_{T2}) > 1300$ MeV/c and $IP_1, IP_2 > 0.5$ mm, the TOS efficiency calculated will be not very different from that of the online muonID. The cuts for Single Muon trigger at L0, HLT1 and HLT2 were summarised in table 6.4.

In order to obtain the mentioned mapping, a $B^+ \to J/\psi K^+$ trigger unbiased[8] data sample has been used. This is the same sample used for the CDLL cut efficiency analysis in previous section. Since the cuts at trigger level are in IP (or IPS, which is highly correlated) and p_T, the initial mapping has been done using these variables. The trigger efficiency has been obtained using the formula $N^{TOS\&\&TIS}/N^{TIS}$ for each of the bins. Figure 6.30 shows the result of the mapping. It can be seen that the efficiency is essentially only dependent on p_T for the muons in $K^0_S \to \mu^+\mu^-$, with IP above 0.5 mm (which was a fiducial cut imposed in section 6.3). The

[8] again, TIS with respect to the three trigger levels.

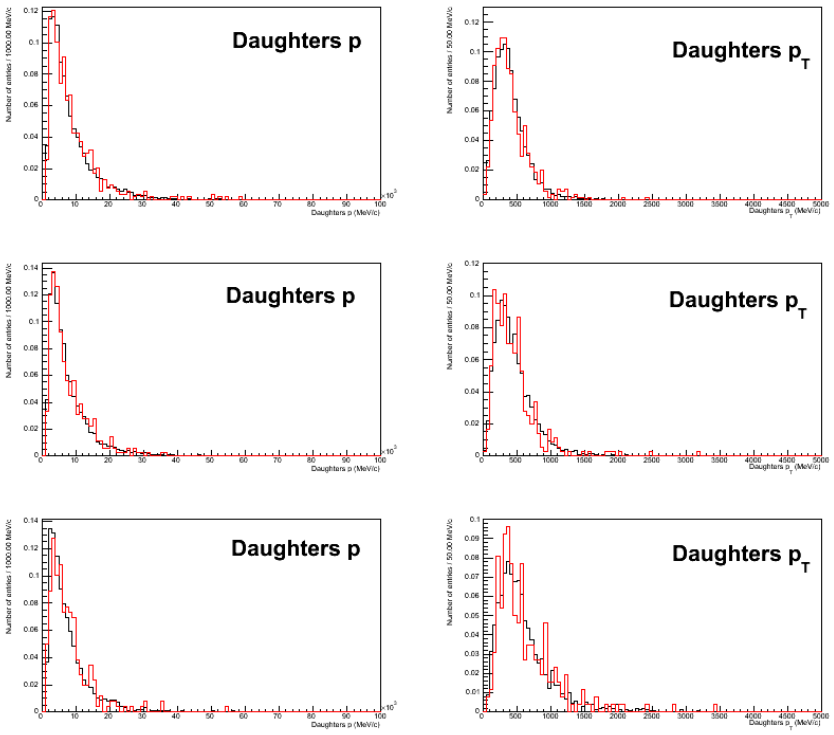

Figure 6.29: p and p_T distributions for three different TIS BDT bins for the MC11 $K^0_S \to \mu^+\mu^-$ (black) and $K^0_S \to \pi^+\pi^-$ (red). The BDT used is the one of sample B. Top plots correspond to bin number 1, medium to bin number 5 and bottom to bin number 10.

apparent effect showed in the plot, giving lower efficiencies at higher IP values, is related to the anti-correlation between IP and p_T [202].

After the results in figure 6.30, the dependency on IP for the TOS efficiency can be dropped. Given that, as explained, the remaining efficiency after the fiducial cut in the maximum p_T of the $K^0_S \to \mu^+\mu^-$ daughters is mainly that of the online muonID, it turns out to be natural to parameterise this efficiency in bins of p and p_T, as done with the offline muonID. The way to obtain this parameterisation is essentially the same seen in section 6.7.2, using $B^+ \to J/\psi K^+$. Two fits to the $J/\psi \to \mu^+\mu^-$ mass spectra have been performed for each bin. One for the muons fulfilling $TOS\&\&TIS\&\&CDLL > -4\&\&IsMuon$ and other for muons with $(!TOS)\&\&TIS\&\&CDLL > -4\&\&IsMuon$. With the obtained number of J/ψ in each case, the efficiency is obtained with $\frac{N(J/\psi)_{TOS\&\&TIS\&\&CDLL>-4\&\&IsMuon}}{(N(J/\psi)_{TOS\&\&TIS\&\&CDLL>-4\&\&IsMuon}+N(J/\psi)_{(!TOS)\&\&TIS\&\&CDLL>-4\&\&IsMuon})}$. Again, the same two fitting models are used. The first included a third order Chebyshev polynomial for the background and a Crystal Ball for the signal. The second fit used a background only exponential model, obtaining the number of J/ψ from the difference in the total number of candidates and the background candidates given by the fit. The central value quoted uses the second fit, while the systematic error is the difference between the efficiencies obtained with both models. The curves are obtained in an independent way for positive and negative muons, to account for any possible asymmetry in the TOS efficiency between both charges. Figure 6.31 shows the curves obtained for the positive case. The explicit results used for the normalisation are shown in table 6.20 (6.21) for positive (negative) muons.

With the obtained curves for the TOS efficiency, the same procedure as for the muonID has been applied. Instead of detaching the muonID and TOS efficiencies, the mapping has been merged by computing $\epsilon^{muonID} \times \epsilon^{TOS}$ as a function of p, p_T. After that, with the idea of

BDT bin	Sample A	Sample B
1	(70.22 ± 0.71) %	(69.93 ± 3.06) %
2	(70.28 ± 0.95) %	(70.41 ± 1.09) %
3	(70.48 ± 0.19) %	(70.27 ± 0.79) %
4	(70.39 ± 2.19) %	(70.52 ± 0.83) %
5	(69.58 ± 2.78) %	(70.46 ± 1.83) %
6	(70.17 ± 3.63) %	(70.55 ± 2.45) %
7	(70.62 ± 4.25) %	(70.35 ± 1.96) %
8	(70.74 ± 3.61) %	(70.90 ± 4.63) %
9	(69.98 ± 6.19) %	(70.86 ± 8.02) %
10	(68.77 ± 3.04) %	(70.20 ± 3.71) %

Table 6.18: MuonID total efficiency in bins of BDT with BDTA and BDTB for $K^0_S \to \mu^+\mu^-$ TIS. The muonID has been evaluated weighting the $K^0_S \to \pi^+\pi^-$ MB data selected after stripping and fiducial cuts using a set of muonID curves obtained with a sample of probe muons coming from J/ψ from B (see text for details).

BDT bin	Sample A	Sample B
1	(75.37 ± 6.70) %	(75.81 ± 7.52) %
2	(76.75 ± 6.83) %	(77.07 ± 7.65) %
3	(79.36 ± 7.06) %	(79.40 ± 7.88) %
4	(78.03 ± 6.94) %	(79.94 ± 7.93) %
5	(78.95 ± 7.02) %	(81.28 ± 8.07) %
6	(77.75 ± 6.91) %	(80.42 ± 7.98) %
7	(79.39 ± 7.06) %	(80.22 ± 7.96) %
8	(79.92 ± 7.11) %	(81.10 ± 8.05) %
9	(79.71 ± 7.09) %	(81.58 ± 8.10) %
10	(80.56 ± 7.16) %	(81.52 ± 8.09) %

Table 6.19: MuonID total efficiency in bins of BDT with BDTA and BDTB for $K^0_S \to \mu^+\mu^-$ TOS. The muonID has been evaluated weighting the $K^0_S \to \pi^+\pi^-$ MB data selected after stripping and fiducial cuts using a set of muonID curves obtained with a sample of probe muons coming from J/ψ from B (see text for details).

obtaining the correct p/p_T spectra, the $K^0_S \to \pi^+\pi^-$ MB data sample has been used, in the corresponding bins of TOS BDTA and BDTB. Again, to remove any remanent background, the K^0_S are restricted to the signal region (492-504 MeV/c^2). The total efficiency has been measured by calculating the total efficiency for each K^0_S candidate (using the $\pi^\pm$ p and p_T and the values in figure 5.10, and tables 6.17, 6.20 and 6.21) and then computing the mean (see equation 6.18).

$$\begin{aligned}\epsilon^{muonID\&\&TOS}_{BDT\ bin\ i} = \Bigg(\sum_j^{N^{cands}_{BDT\ bin\ i}} & \epsilon_{muonID}(p_{j1}, p_{Tj1}) \times \epsilon_{muonID}(p_{j2}, p_{Tj2}) \times \\ & \times [\ \epsilon^+_{TOS}(p_{j1}, p_{Tj1}) + \epsilon^-_{TOS}(p_{j2}, p_{Tj2}) - \\ & - \epsilon^+_{TOS}(p_{j1}, p_{Tj1}) \times \epsilon^-_{TOS}(p_{j2}, p_{Tj2})\ ] \Bigg) / N^{cands}_{BDT\ bin\ i}\end{aligned} \tag{6.18}$$

As in the muonID case, total lower and upper errors are also calculated in a similar way, changing for every p/p_T bin the initial value by the value $\pm$ the error. It must be also reminded that if a $\pi^\pm$ from the K^0_S has a p or p_T out of the defined binning the closest bin is used.

A full complete systematic error is now added repeating the procedure just explained in MC.

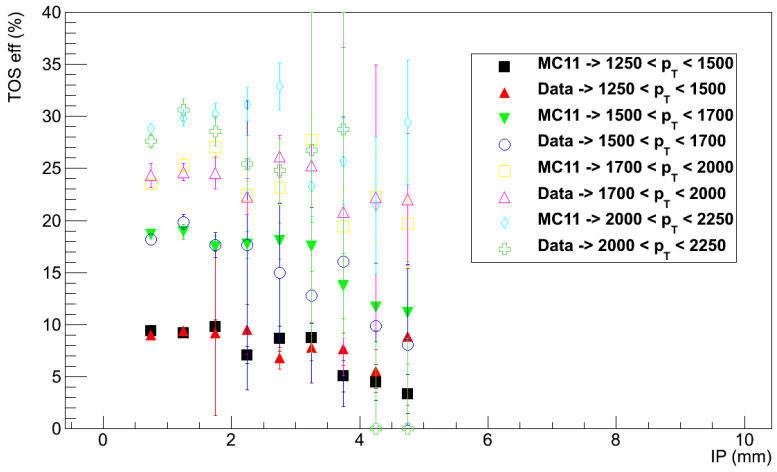

Figure 6.30: $\epsilon_{SingleMuon\ TOS}$ vs. IP in bins of p_T for a set of trigger unbiased muons from $B^+ \to J/\psi K^+$ (data and MC11) fulfilling IsMuon and CDLL>-4. The efficiency in data is calculated with $N^{TIS\&\&TOS}/N^{TIS}$, while in MC by counting the number of events before and after the cuts. The p_T bins are defined in MeV/c.

	$0.05 < p_T < 1.3$	$1.3 < p_T < 1.7$	$1.7 < p_T < 3.0$	$3.0 < p_T < 5.0$
$3 < p < 10$	(0.02 ± 0.01) %	(5.96 ± 0.32) %	(9.79 ± 0.51) %	(1.59 ± 2.90) %
$10 < p < 21$	(0.09 ± 0.02) %	(16.07 ± 0.37) %	(28.01 ± 0.86) %	(29.86 ± 1.36) %
$21 < p < 30$	(0.04 ± 0.03) %	(18.93 ± 1.22) %	(32.60 ± 0.62) %	(36.91 ± 0.89) %
$30 < p < 40$	(0.05 ± 0.05) %	(19.17 ± 0.80) %	(32.29 ± 0.79) %	(39.36 ± 1.03) %
$40 < p < 60$	(0.13 ± 0.09) %	(18.82 ± 0.83) %	(34.15 ± 1.29) %	(37.48 ± 1.00) %

Table 6.20: $\epsilon_{SingleMuon\ TOS}$ vs. p/p_T with respect to IsMuon and CDLL>-4 obtained from a set of trigger unbiased positive muons from $B^+ \to J/\psi K^+$. All the bins are defined in GeV/c. The errors quoted include a statistical and systematical contribution.

The same weighting is applied to the MC11 $K^0_S \to \mu^+\mu^-$ after the TOS BDT cut of the first BDT bin (given the low statistics after the fiducial cut in $max(p_{T1}, p_{T2}) > 1300$ MeV/c, is not possible to separate in all the bins of BDT). The result obtained in MC is then compared to the one obtained using the MC Truth information (simply by counting number of events before and after the cuts). Since the muonID and TOS efficiencies are known to be very similar in data and MC (for TOS efficiency, see figure 6.30), this comparison helps to quantify the inaccuracy due to the use of the curves of ϵ^{muonID} and ϵ^{TOS}. Therefore, the difference between the results with both methods in MC is then assigned as a systematic error in data. As in the muonID case, this error will become dominant with respect to the one arising from the errors in $\epsilon^{muonID} \times \epsilon^{TOS}$ tables, and it will be the only one considered.

The results obtained are shown in table 6.22. Finally, an indication of the TOS efficiency only in bins of BDT is shown in table 6.23. This table is only informative, and is computed as the ratio between tables 6.22 and 6.19.

6.7.4 Mass binning

Inside each BDT bin, the invariant mass distribution of the $K^0_S \to \pi^+\pi^-$ events could be used to predict the $K^0_S \to \mu^+\mu^-$ invariant mass PDF. However, here the mass of the daughters accounts for a sizeable fraction of the total mass. As a side effect, the mass resolution (and therefore the mass PDF) differs significantly between $K^0_S \to \mu^+\mu^-$ and $K^0_S \to \pi^+\pi^-$. Figure 6.32 shows

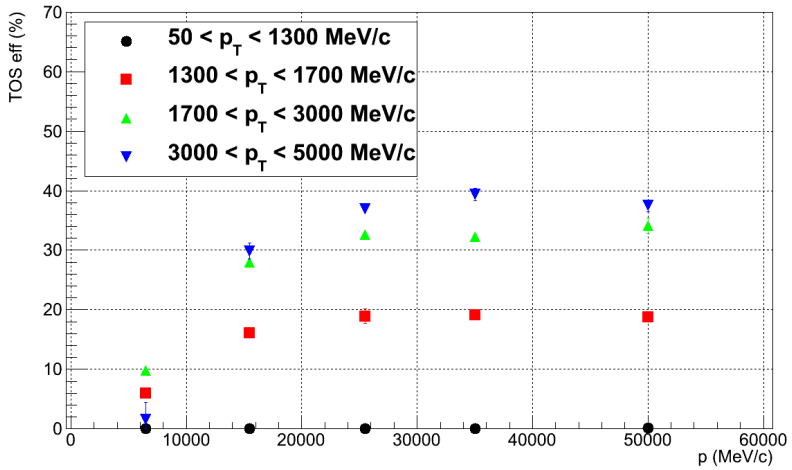

Figure 6.31: $\epsilon_{SingleMuon\ TOS}$ vs. p/p_T for a set of trigger unbiased positive muons from data $B^+ \to J/\psi K^+$ fulfilling IsMuon and CDLL>-4. The errors showed include a statistical and systematical contribution. The efficiency is calculated with $N^{TIS\&\&TOS}/N^{TIS}$.

	$0.1 < p_T < 1.3$	$1.3 < p_T < 1.7$	$1.7 < p_T < 3.0$	$3.0 < p_T < 5.0$
$3 < p < 10$	(0.00 ± 0.01) %	(5.54 ± 0.28) %	(10.94 ± 0.93) %	(1.69 ± 2.93) %
$10 < p < 21$	(0.08 ± 0.02) %	(17.19 ± 0.36) %	(28.21 ± 0.39) %	(28.98 ± 0.99) %
$21 < p < 30$	(0.06 ± 0.03) %	(19.31 ± 0.76) %	(31.69 ± 0.81) %	(36.79 ± 0.97) %
$30 < p < 40$	(0.14 ± 0.08) %	(20.79 ± 2.89) %	(33.96 ± 0.80) %	(40.39 ± 0.91) %
$40 < p < 60$	(0.07 ± 0.08) %	(18.39 ± 14.45) %	(32.32 ± 0.70) %	(39.04 ± 0.95) %

Table 6.21: $\epsilon_{SingleMuon\ TOS}$ vs. p/p_T with respect to IsMuon and CDLL>-4 obtained from a set of trigger unbiased negative muons from $B^+ \to J/\psi K^+$. All the bins are defined in GeV/c. The errors quoted include a statistical and systematical contribution.

the $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu^+\mu^-$ mass peaks in MC simulation. For this analysis, it has been decided to use a single mass bin, to not rely on a precise description of the mass PDF. As already mentioned, the mass window is chosen to be 492–504 MeV/c^2. The efficiency of such a mass cut is between 85% and 95% depending on the channel ($K^0_S \to \pi^+\pi^-$ or $K^0_S \to \mu^+\mu^-$), BDT bin, and assumed mass resolution difference between data and MC. An efficiency ratio of $\epsilon^M_{\pi\pi}/\epsilon^M_{\mu\mu} = 1.08 \pm 0.04$ is added as a global factor. Table 6.24 shows the ratio of the mass window cut efficiencies between $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu^+\mu^-$ as a function of the BDT bin, for BDTA and BDTB in TIS and TOS categories. The ratio is computed using MC11 samples. A systematic error is also obtained by smearing the mass resolution in MC as $\sigma_{data}/\sigma_{MC} = 1.1 \pm 0.1$ (where the resolutions are measured by simple Gaussian fits, and taken in different BDT bins). This systematic also covers the small variations of the ratio of efficiencies across BDT bins and the four BDT types.

6.7.5 $K^0_S \to \pi^+\pi^-$ TIS yields and normalisation factors for TIS analysis

To get the final normalisation TIS factor, the only remaining information needed is the number of observed $K^0_S \to \pi^+\pi^-$ (TIS) events in each BDT bin. This number is calculated by counting events, without performing any mass fit as the background is negligible even in the lower BDT bins (see figure 6.33). Table 6.25 shows the number of observed $K^0_S \to \pi^+\pi^-$ events per BDT bin in each of the samples. The assumption of the ratio of TIS efficiencies being 1, has been validated at the 18% level using MC offline reconstructed events. This 18% will be added as a

BDT bin	Sample A	Sample B
1	(17.13 ± 2.74) %	(16.02 ± 2.82) %
2	(16.88 ± 2.70) %	(15.58 ± 2.74) %
3	(17.24 ± 2.76) %	(16.26 ± 2.86) %
4	(16.63 ± 2.66) %	(16.33 ± 2.87) %
5	(16.11 ± 2.58) %	(16.78 ± 2.95) %
6	(15.47 ± 2.48) %	(16.33 ± 2.88) %
7	(15.62 ± 2.50) %	(15.85 ± 2.79) %
8	(15.57 ± 2.49) %	(15.83 ± 2.79) %
9	(15.49 ± 2.48) %	(16.23 ± 2.86) %
10	(14.78 ± 2.37) %	(15.23 ± 2.68) %

Table 6.22: Trigger and muonID total efficiency in bins of BDT with BDTA and BDTB for $K^0_S \to \mu^+\mu^-$ TOS. The muonID and trigger efficiencies have been evaluated weighting the $K^0_S \to \pi^+\pi^-$ MB data selected after stripping and fiducial cuts using a set of muonID and TOS efficiency curves obtained with a sample of probe muons coming from J/ψ from B (see text for details).

BDT bin	Sample A	Sample B
1	22.76 %	21.16 %
2	22.04 %	20.27 %
3	21.77 %	20.47 %
4	21.43 %	20.50 %
5	20.42 %	20.61 %
6	19.98 %	20.47 %
7	19.83 %	19.79 %
8	19.52 %	19.57 %
9	19.53 %	20.02 %
10	18.37 %	18.67 %

Table 6.23: Trigger efficiency in bins of BDT with BDTA and BDTB for $K^0_S \to \mu^+\mu^-$ TOS. The trigger efficiencies are computed as the ratio between tables 6.22 and 6.19

systematic, affecting simultaneously all bins.

When calculating the normalisation factor, the fact that the $K^0_S \to \pi^+\pi^-$ selection is prescaled by a factor of 1000 also has to be taken into account. Using $\mathcal{B}(K^0_S \to \pi^+\pi^-)$ from [5], the normalisation factors per BDT bin shown in table 6.26 are obtained.

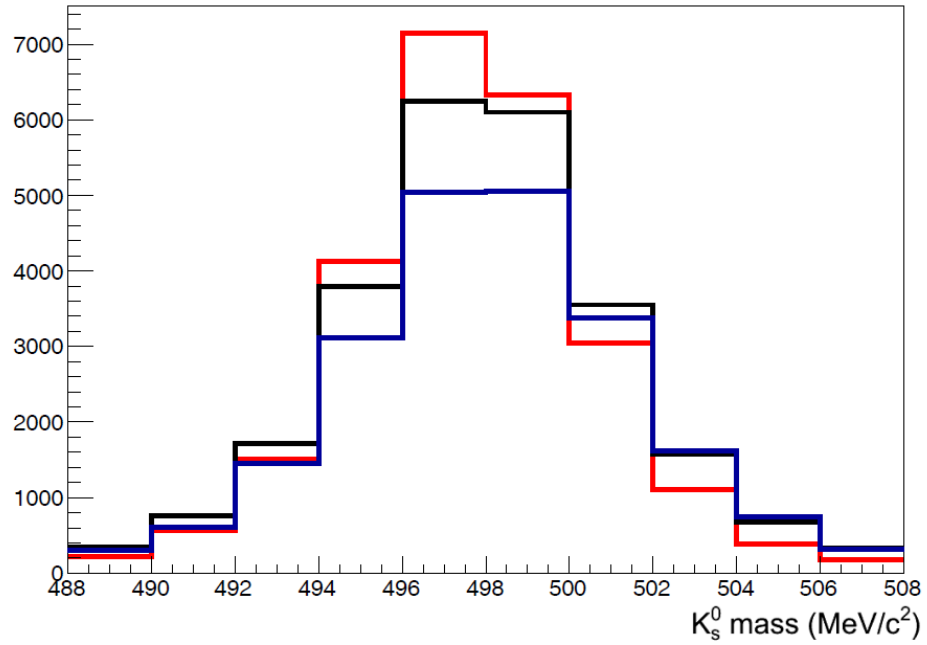

Figure 6.32: Invariant mass distribution of the $K^0_S \to \pi^+\pi^-$ (red) and $K^0_S \to \mu^+\mu^-$ (blue) in MC simulation. Black: data from MB stream.

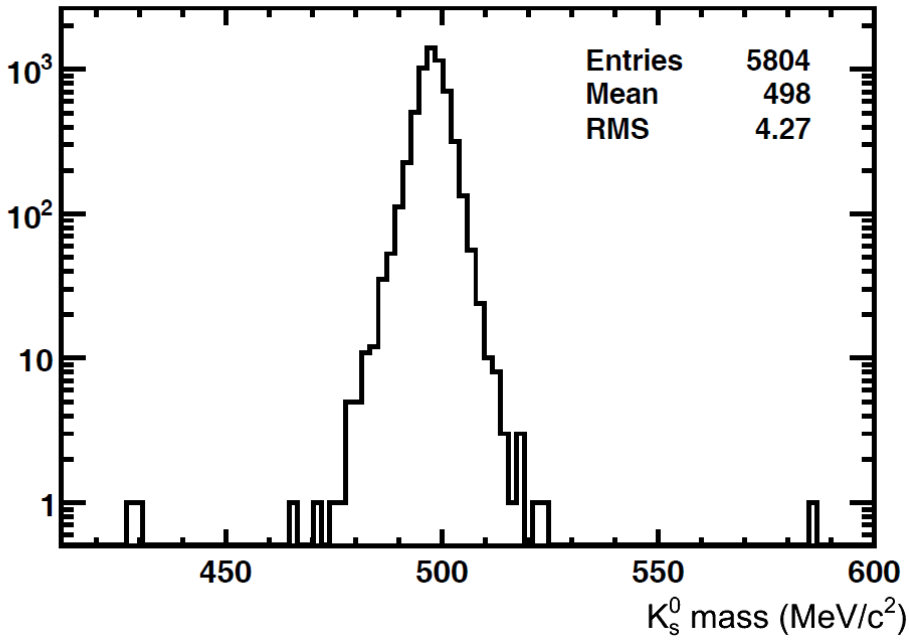

Figure 6.33: Invariant mass distribution of the $K^0_S \to \pi^+\pi^-$ candidates in the lowest BDTA TIS bin, in logarithmic scale.

BDT	TIS		TOS	
bin	Sample A	Sample B	Sample A	Sample B
1	1.093 ± 0.026	1.072 ± 0.028	1.179 ± 0.061	0.850 ± 0.248
2	1.061 ± 0.024	1.052 ± 0.027	1.235 ± 0.092	0.922 ± 0.212
3	1.076 ± 0.021	1.059 ± 0.026	1.273 ± 0.126	1.156 ± 0.075
4	1.043 ± 0.020	1.052 ± 0.023	1.077 ± 0.056	1.146 ± 0.064
5	1.034 ± 0.021	1.035 ± 0.024	0.558 ± 0.396	1.079 ± 0.047
6	1.040 ± 0.016	1.039 ± 0.018	1.129 ± 0.069	1.176 ± 0.078
7	1.043 ± 0.014	1.038 ± 0.016	1.130 ± 0.080	0.891 ± 0.265
8	1.006 ± 0.020	1.036 ± 0.015	1.057 ± 0.042	1.138 ± 0.074
9	1.042 ± 0.014	1.035 ± 0.015	1.109 ± 0.051	1.217 ± 0.107
10	1.055 ± 0.018	1.037 ± 0.014	1.049 ± 0.081	1.150 ± 0.093

Table 6.24: Ratio of the mass window cut efficiencies between $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu^+\mu^-$ as a function of the BDT bin, for BDTA and BDTB in TIS and TOS categories. The ratio is computed using MC11 samples.

BDT bin	Sample A	Sample B
1	8125	8807
2	8629	9085
3	9103	9234
4	9579	9394
5	9701	9559
6	10111	9795
7	9493	9956
8	8586	9998
9	7594	10043
10	5124	10053

Table 6.25: Observed $K^0_S \to \pi^+\pi^-$ (TIS) yields in samples A (left) and B (right), in bins of BDT. The numbers corresponds to candidates in the search window.

BDT bin	α^A_{TIS}	α^B_{TIS}
1	8.12 ± 0.39	7.40 ± 0.48
2	8.68 ± 0.53	8.21 ± 0.52
3	8.11 ± 0.55	7.46 ± 0.46
4	8.26 ± 0.68	7.41 ± 0.50
5	8.12 ± 0.81	8.65 ± 0.77
6	6.63 ± 0.60	7.80 ± 1.10
7	8.41 ± 0.97	8.18 ± 0.75
8	10.30 ± 1.70	7.55 ± 0.89
9	11.50 ± 2.00	7.60 ± 1.20
10	16.20 ± 1.50	7.98 ± 0.81

Table 6.26: Normalisation factors (in units of 10^{-8}) for the sample A (left) and B (right), in bins of BDT TIS. In addition to those, there is a 18% error coming from the validation in MC of the ratio of TIS efficiencies, that affects coherently all the BDT bins. The error on $\mathcal{B}(K^0_S \to \pi^+\pi^-)$ also affects coherently all the bins of the TIS sample, but is negligible.

6.7.6 $K^0_S \to \pi^+\pi^-$ MB yields and normalisation factors for TOS analysis

BDT bin	Sample A	Sample B
1	1310	1694
2	857	1284
3	634	929
4	573	875
5	587	729
6	574	723
7	578	640
8	705	611
9	774	581
10	2306	544

Table 6.27: Observed $K^0_S \to \pi^+\pi^-$ (MB) yields in bins of BDT TOS A and BDT TOS B. The numbers correspond to candidates in the search window.

To normalise $K^0_S \to \mu^+\mu^-$ TOS to $K^0_S \to \pi^+\pi^-$, the trigger efficiency of $K^0_S \to \pi^+\pi^-$ has to be known. One possibility would be using the formalism of [179], as in the case of $B^0 \to \mu^+\mu^-$ analysis (section 5.9.5). However, the current sample contains only 11 $K^0_S \to \pi^+\pi^-$ TIS&TOS for sample A and 13 for sample B, which would need to be binned in p_T in order to get an estimation of the trigger efficiency. Another option could be taking $K^0_S \to \pi^+\pi^-$ from the MB sample, but this would imply that the average prescale factor s^{MB} should also be under control. Instead of choosing one or the other, both can be used in order to get the total amount of $K^0_S \to \pi^+\pi^-$ that would have been selected if there were no prescale factors at any level. Following the definition of the trigger efficiency, and considering the prescale factor of the $K^0_S \to \pi^+\pi^-$ stripping selection (1000):

$$N^{SEL}_{K^0_S\to\pi^+\pi^-} = 1000 \times \frac{N^{TRIG}_{K^0_S\to\pi^+\pi^-}}{\epsilon^{TRIG}} \tag{6.19}$$

For each TCK, the trigger efficiency can be obtained from the MB sample. Therefore:

$$N^{SEL}_{K^0_S\to\pi^+\pi^-} = 1000 \times \sum_i^{TCK} N^{TRIG,i}_{K^0_S\to\pi^+\pi^-} \times \frac{N^{MB,i}_{K^0_S\to\pi^+\pi^-}}{N^{MB\&TRIG,i}_{K^0_S\to\pi^+\pi^-}} \tag{6.20}$$

Then, $\frac{N^{MB}_{K^0_S\to\pi^+\pi^-}}{N^{SEL}_{K^0_S\to\pi^+\pi^-}}$ gives an averaged prescale factor s^{MB} for the MB events or, alternatively, $\frac{N^{TRIG}_{K^0_S\to\pi^+\pi^-}}{N^{SEL}_{K^0_S\to\pi^+\pi^-}}$ averaged trigger efficiencies. With this, an average prescale factor of $s^{MB} = (2.7 \pm 0.3) \times 10^{-6}$ and TIS efficiency (for events that pass the GECs too) of $\sim 0.15\%$ can be obtained.

The just obtained MB prescale factor , s^{MB}, can be compared with a direct calculation based on the prescale factors of the MB no–bias line[9]. This line selects random events with a given rate, and so that 70% of them correspond to beam–beam crossing (bb). Using 14.9 MHz as input rate of bb, the average prescale factor is then given by:

$$\sum_i^{TCK} Lumi_i \times 0.7 \times \frac{Rate^{MB,i}}{14.9\text{MHz}} \tag{6.21}$$

where $Lumi_i$ is the fraction of integrated luminosity corresponding to each TCK (which can be

[9]The trigger line which selects events randomly rather than based on any physics decision.

extracted from table 3.1) and $Rate^{MB,i}$ the MB no–bias rate per TCK.

Using MB no–bias rates, $Rate^{MB,i}$, of 11 Hz and 97 Hz (depending on the TCK, but in both cases triggering 70% of the times in bb), an average prescale of $s^{MB} \sim 2.0 \times 10^{-6}$ is measured. The difference between this and the previous estimation is assigned as a systematic error, thus obtaining $s^{MB} = (2.70 \pm 0.76) \times 10^{-6}$.

Also, the obtained TIS efficiency can be compared with $\frac{N^{TIS\&TOS}}{N^{TOS}}$ in the $K^0_S \to \pi^+\pi^-$ misID sample, which following [179] should also yield an estimation of the TIS efficiency. Doing so, a value of $\epsilon_{TIS} \sim 0.11$ % is obtained. However, this number is quite dependent on the requirement or not of *nShared* for the muons, varying between 0.08 % and 0.38 %.

Finally, it also needs to be taken into account that, concerning the TOS category, 44% of the data corresponds to sample A, and 56% to sample B. The total number of $K^0_S \to \pi^+\pi^-$ candidates in MB sample and in bins of BDT TOS is given in table 6.27. The final TOS normalisation factors are given in table 6.28. Notice that the differences seen in the normalisation factors between the different bins in table 6.28 (and also in table 6.26, in the TIS case) arise both from the BDT shape seen in figure 6.17 and from the statistical fluctuations in the ratios of efficiencies as obtained from MC.

As a reference, the total normalisation factor, including the 4 samples (TIS A, TIS B, TOS A and TOS B) is $\approx 5.6 \times 10^{-10}$.

BDT bin	α^A_{TOS}	α^B_{TOS}
1	1.94 ± 0.70	0.92 ± 0.24
2	1.84 ± 0.65	2.40 ± 1.00
3	3.40 ± 1.40	1.61 ± 0.54
4	5.50 ± 2.20	1.24 ± 0.48
5	2.50 ± 1.00	2.50 ± 1.10
6	3.90 ± 1.60	4.60 ± 3.20
7	7.80 ± 4.80	2.01 ± 0.69
8	2.41 ± 0.97	4.20 ± 1.90
9	2.30 ± 0.99	4.20 ± 3.30
10	0.94 ± 0.22	7.10 ± 3.20

Table 6.28: Normalisation factors (in units of 10^{-8}) for the sample A (left) and B (right), in bins of BDT TOS. The errors on this table are uncorrelated. In addition to those, there is a 28% error coming from the effective prescale factor, that affects coherently all the BDT bins. The error on $\mathcal{B}(K^0_S \to \pi^+\pi^-)$ also affects coherently all the bins of the TOS sample, but is negligible.

6.8 Background level

The expectation for the sum of the combinatorial background and the doubly-misidentified events is obtained by a fit to the data mass sidebands. The model for the combinatorial background is an exponential. The model for the doubly-misidentified background is a power law, as described in section 6.6.2. A linear combination of these two models is then used.

The mass sidebands considered in the fit are [470,492] MeV/c^2 and [504, 600] MeV/c^2. The signal window [492, 504] MeV/c^2 is not used, and it was blinded during the analysis.

This fit is performed separately for each bin of BDT, for each BDT type (A,B), and for each trigger category (TIS,TOS). Figures 6.34, 6.35, 6.36 and 6.37 show the fits for the combination of the two trigger categories and sample types.

The results per bin of BDT are gathered in tables 6.29 and 6.30 for trigger type TIS and TOS, respectively. The left column of both tables holds the expected number of background events in the search window [492, 504] MeV/c^2.

In order to estimate a systematical uncertainty on the background expectation, other fit models are used to extract those numbers. An alternative model consists of the sum of a power law and a linear function, fitted on the same sidebands as the base model, [470,492] MeV/c^2 and [504,600] MeV/c^2. The other model is the same as the base model, but in a slightly reduced lower mass sideband [475,492] MeV/c^2 keeping the upper sideband to [504,600] MeV/c^2. The results of the alternatives fits are also gathered in tables 6.29 and 6.30, in central and right columns. In the few cases where the difference in the expected background events was found larger than their combined uncertainties, a systematic uncertainty is assigned. Its value is estimated with the relative difference between the two model expectations, and it can be found in the last column of the mentioned tables when applicable.

Another test consists in comparing the number of events observed in regions around the signal windows with the expected number of event as obtained from a fit. For the test to be valid, the fit should avoid the region under scrutiny as well as the blinded region. Hence, for each bin, a fit is performed using [470,480]∪[504,600] MeV/c^2 and the expected number of events in the region [480,492] MeV/c^2 computed. As this region is not blinded, the number of events in it can actually be counted. A similar procedure is performed for the region above the signal mass window: for each bin a fit is performed on the [470,492]∪[516,600] MeV/c^2 sidebands, from which the expected number of events in the [504,516] MeV/c^2 region is extracted. These expected numbers, together with the actual number of event in the same mass windows are reported in tables 6.31 and 6.32. For each bin and each mass region, the expected number is compared to the actual number of events. Feldman-Cousins uncertainties at 68% confidence level are assumed for the observed number of events. In case there are discrepancies larger than one σ between the observation and the expectations a systematical uncertainty is computed. This is done so that the number of events expected in the search window is bound by the observed events in the side-windows.

Both the reported systematics are finally added in quadrature to the statistical uncertainty for the affected bins.

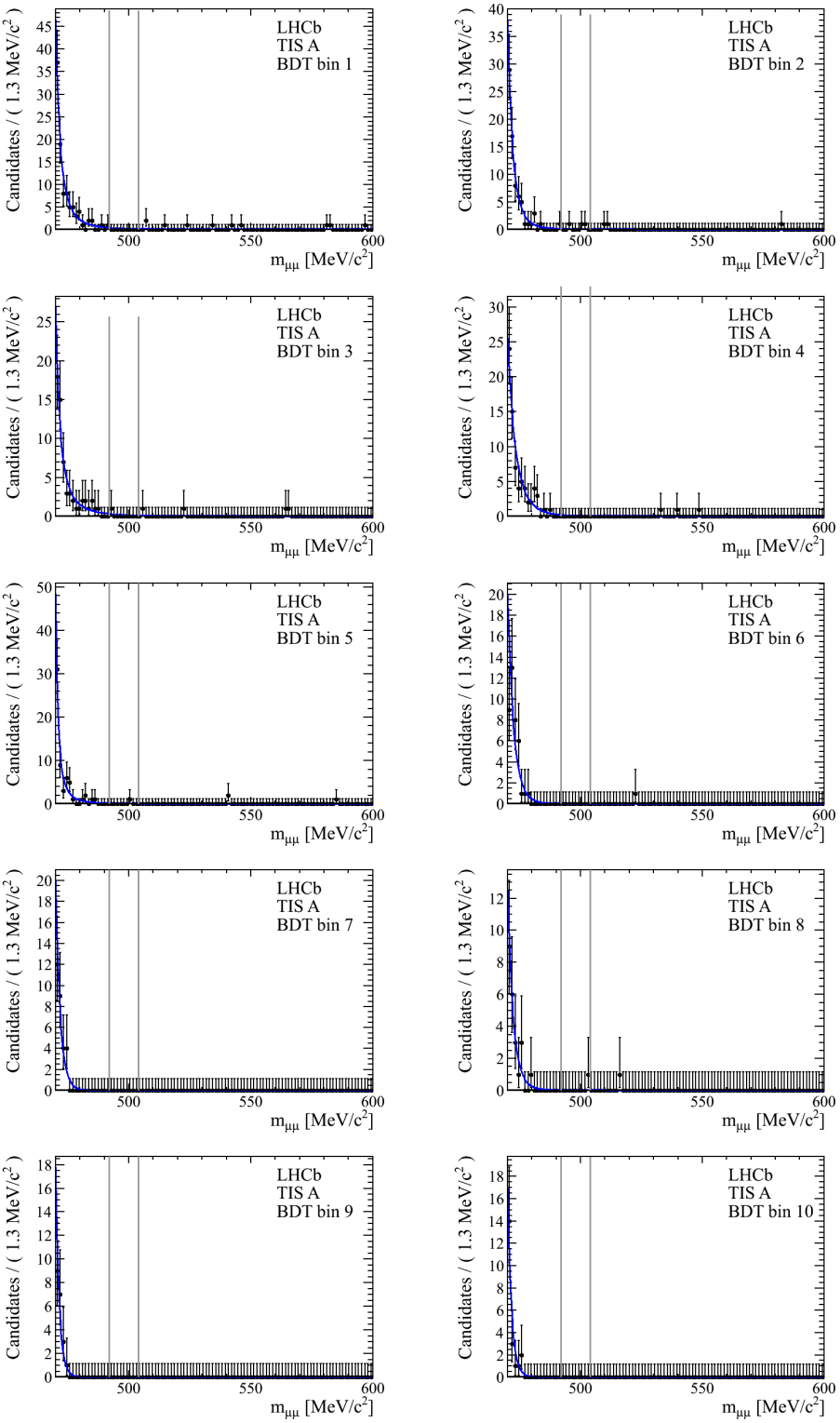

Figure 6.34: Background mass fits for TIS events in sample A. The vertical lines delimit the search window. The bins are ordered according to left-to-right script: first bin on top row on the left and last bin on last row on the right.

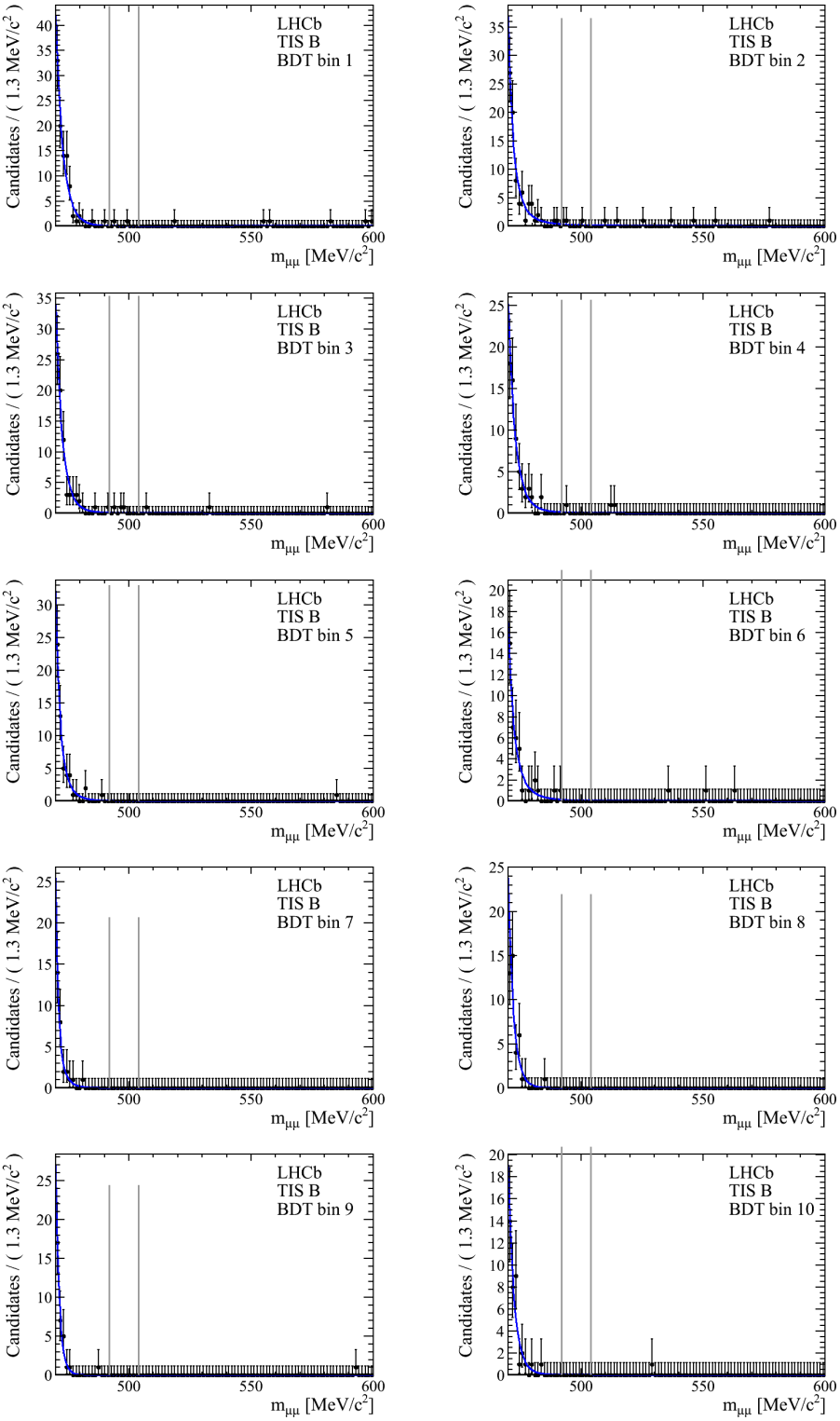

Figure 6.35: Background mass fits for TIS events in sample B. The vertical lines delimit the search window. The bins are ordered according to left-to-right script: first bin on top row on the left and last bin on last row on the right.

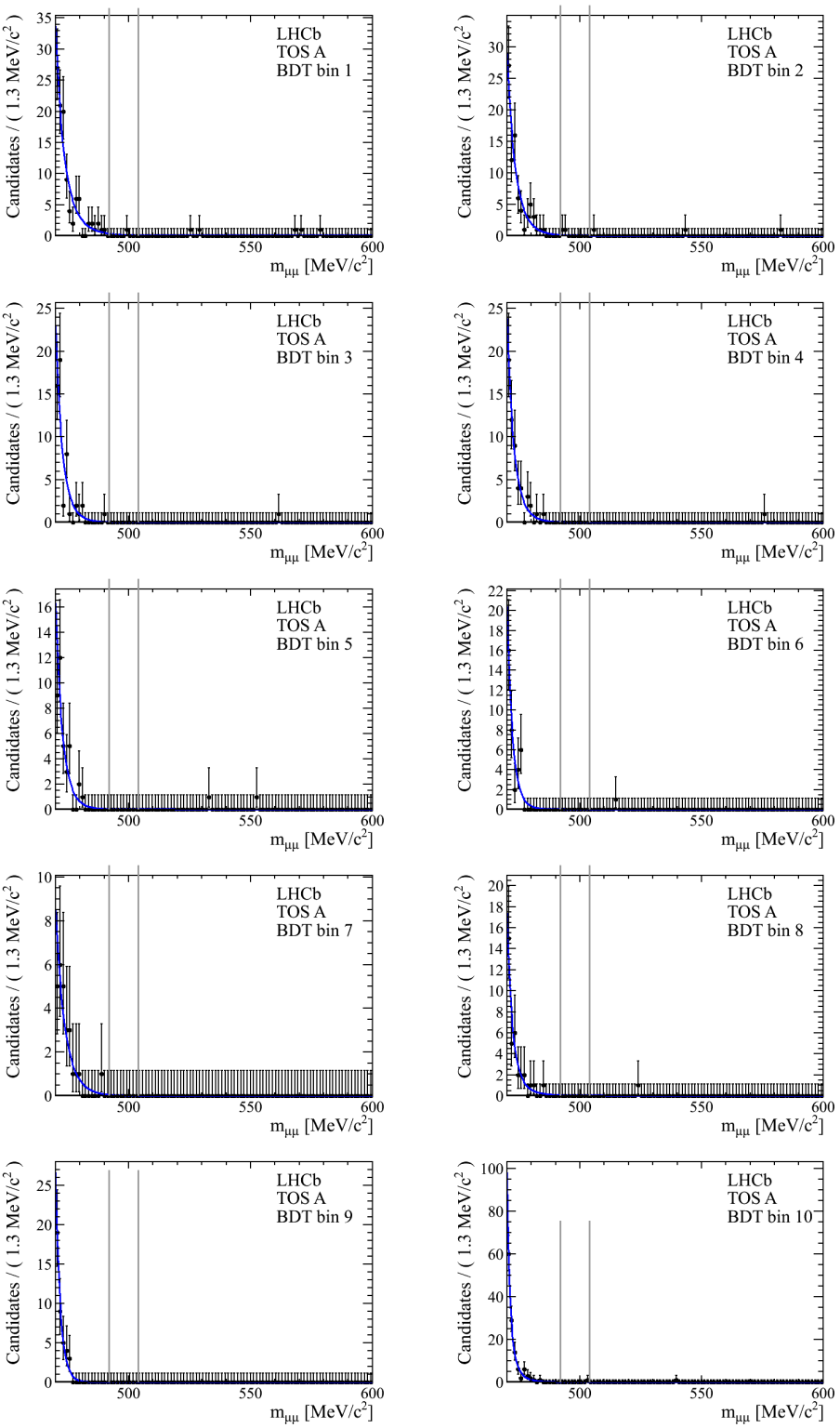

Figure 6.36: Background mass fits for TOS events in sample A. The vertical lines delimit the search window. The bins are ordered according to left-to-right script: first bin on top row on the left and last bin on last row on the right.

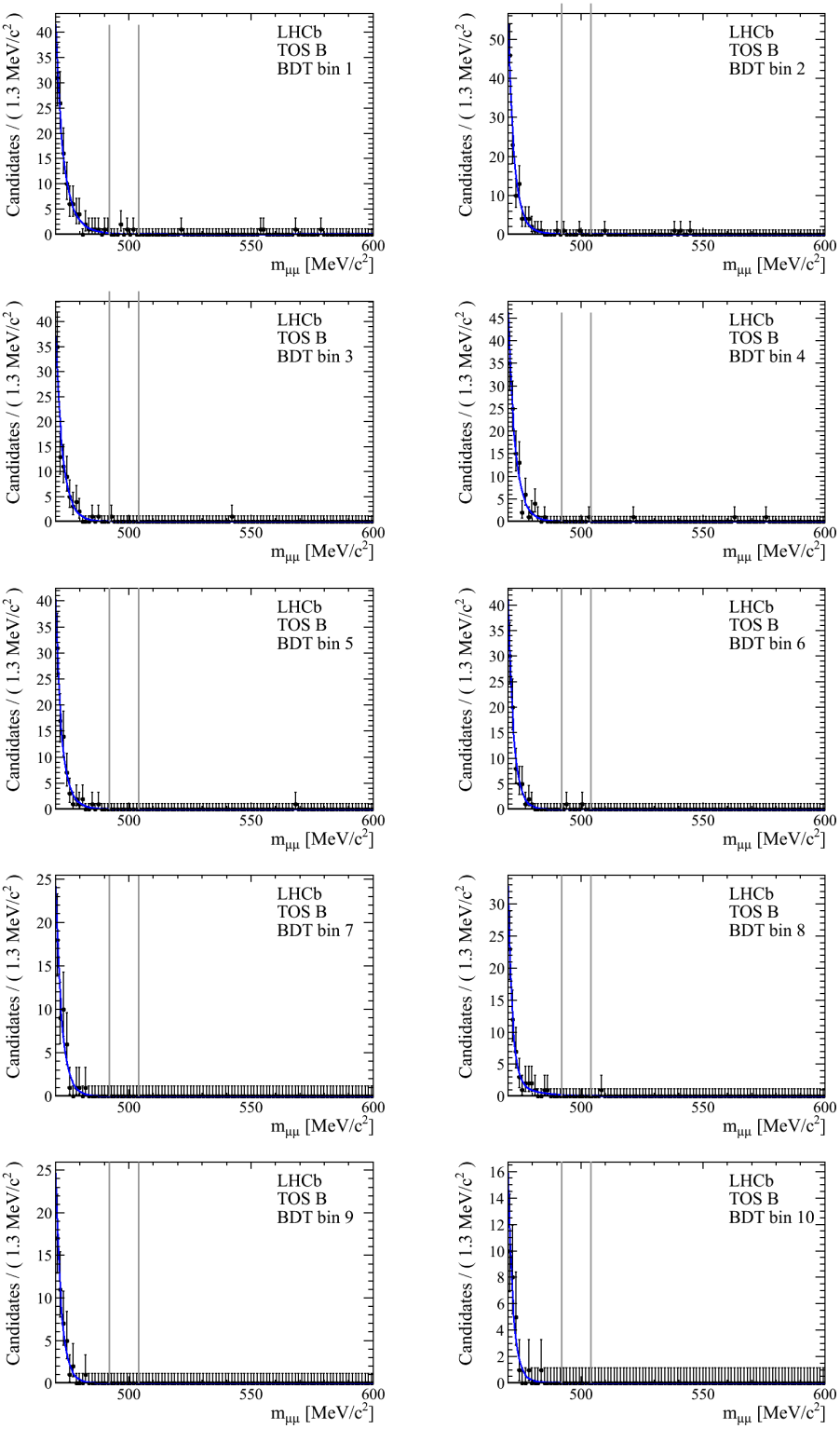

Figure 6.37: Background mass fits for TOS events in sample B. The vertical lines delimit the search window. The bins are ordered according to left-to-right script: first bin on top row on the left and last bin on last row on the right.

sample	bin	base model	reduced lower sideband	linear	added syst.
A	bin 1	$3.2^{+1.4}_{-3.2}$	$3.4^{+1.2}_{-1.4}$	$3.2^{+1.3}_{-1.0}$	—
A	bin 2	$1.29^{+1.16}_{-0.66}$	$1.46^{+0.86}_{-0.62}$	$1.23^{+0.81}_{-0.59}$	—
A	bin 3	$2.26^{+1.2}_{-0.87}$	$2.93^{+1.08}_{-0.61}$	$2.22^{+1.07}_{-0.87}$	—
A	bin 4	$1.44^{+1.24}_{-0.75}$	$1.39^{+1.05}_{-0.66}$	$1.2^{+1.33}_{-0.58}$	—
A	bin 5	$1.14^{+0.73}_{-0.54}$	$1.1^{+0.92}_{-0.74}$	$1.17^{+0.72}_{-0.51}$	—
A	bin 6	$0.25^{+0.41}_{-0.25}$	$0.31^{+0.29}_{-0.18}$	$0.145^{+0.167}_{-0.091}$	—
A	bin 7	$0.00139^{+0.00028}_{-0.00111}$	$8.4e-05^{+0.000118}_{-6.7e-05}$	$0.0014^{+0.0454}_{-0.0012}$	—
A	bin 8	$0.31^{+0.48}_{-0.31}$	$0.53^{+0.47}_{-0.29}$	$0.135^{+0.17}_{-0.089}$	—
A	bin 9	$0.00028^{+0.00645}_{-0.00018}$	$1.9e-05^{+0.121087}_{-1.9e-05}$	$0.000183^{+6.9e-05}_{-6.6e-05}$	—
A	bin 10	$0.000193^{+4.6e-05}_{-0.000173}$	$3.7e-07^{+1.6e-06}_{-2.2e-07}$	$0.00038^{+0.00176}_{-0.00026}$	—
B	bin 1	$0.6^{+0.31}_{-0.18}$	$0.93^{+0.67}_{-0.43}$	$0.84^{+0.32}_{-0.25}$	—
B	bin 2	$2.75^{+1.36}_{-0.64}$	$2.47^{+1.35}_{-0.96}$	$2.6^{+1.3}_{-1.1}$	—
B	bin 3	$1.15^{+0.99}_{-0.64}$	$1.11^{+1.01}_{-0.58}$	$0.73^{+0.74}_{-0.35}$	—
B	bin 4	$1.08^{+0.9}_{-0.58}$	$1.14^{+0.99}_{-0.64}$	$0.41^{+0.26}_{-0.18}$	—
B	bin 5	$0.35^{+0.44}_{-0.35}$	$0.32^{+0.48}_{-0.32}$	$0.39^{+0.44}_{-0.39}$	—
B	bin 6	$1.07^{+0.99}_{-0.57}$	$1.61^{+0.87}_{-0.44}$	$0.98^{+0.76}_{-0.49}$	—
B	bin 7	$0.102^{+0.342}_{-0.096}$	$0.008^{+0.2419}_{0.008}$	$0.11^{+0.28}_{-0.1}$	—
B	bin 8	$0.081^{+0.385}_{-0.081}$	$0.083^{+0.495}_{-0.078}$	$0.0149^{+0.0275}_{-0.0099}$	—
B	bin 9	$0.23^{+0.37}_{-0.23}$	$0.2^{+0.27}_{-0.11}$	$0.22^{+0.22}_{-0.13}$	—
B	bin 10	$0.37^{+0.63}_{-0.37}$	$0.5^{+0.81}_{-0.5}$	$0.16^{+0.15}_{-0.1}$	—

Table 6.29: Background expectations in the signal mass window [492,504] MeV/c^2, using the base model and two alternative models (sum of a power-law and linear regression and base model with reduced lower sideband), for the TIS trigger category. In case the asymmetric uncertainties were not available, the symmetric ones are shown.

sample	bin	base model	reduced lower sideband	linear	added syst.
A	bin 1	$2.05^{+1.31}_{-0.91}$	$4.24^{+1.04}_{-0.81}$	$2.05^{+1.3}_{-0.83}$	106%
A	bin 2	$0.86^{+0.73}_{-0.39}$	$1.29^{+1.11}_{-0.67}$	$0.83^{+0.41}_{-0.33}$	—
A	bin 3	$0.23^{+0.4}_{-0.23}$	$0.53^{+0.69}_{-0.34}$	$0.26^{+0.39}_{-0.15}$	—
A	bin 4	$0.23^{+0.5}_{-0.23}$	$0.16^{+0.46}_{-0.1}$	$0.27^{+0.5}_{-0.27}$	—
A	bin 5	$0.35^{+0.53}_{-0.35}$	$0.31^{+0.56}_{-0.31}$	$0.32^{+0.15}_{-0.17}$	—
A	bin 6	$0.28^{+0.45}_{-0.28}$	$0.24^{+0.38}_{-0.24}$	$0.132^{+0.167}_{-0.087}$	—
A	bin 7	$0.21^{+0.36}_{-0.14}$	$0.32^{+0.5}_{-0.29}$	$0.21^{+0.36}_{-0.14}$	—
A	bin 8	$0.59^{+0.8}_{-0.59}$	$0.53^{+0.93}_{-0.53}$	$0.74^{+0.48}_{-0.62}$	—
A	bin 9	$0.00268^{+0.00045}_{-0.00198}$	$8.6e-05^{+6.9e-05}_{-4.9e-05}$	$0.0027^{+0.0205}_{-0.0021}$	96%
A	bin 10	$0.68^{+0.69}_{-0.43}$	$0.29^{+0.54}_{-0.29}$	$0.65^{+0.71}_{-0.4}$	—
B	bin 1	$1.66^{+1.1}_{-0.78}$	$1.64^{+0.49}_{-0.67}$	$1.68^{+1.08}_{-0.71}$	—
B	bin 2	$1.51^{+1.14}_{-0.75}$	$1.46^{+1.13}_{-0.73}$	$0.91^{+0.8}_{-0.41}$	—
B	bin 3	$0.39^{+0.8}_{-0.39}$	$0.66^{+1.04}_{-0.66}$	$0.37^{+0.57}_{-0.37}$	—
B	bin 4	$0.46^{+0.55}_{-0.21}$	$0.57^{+0.62}_{-0.31}$	$0.49^{+0.47}_{-0.18}$	—
B	bin 5	$0.3^{+0.45}_{-0.2}$	$0.41^{+0.59}_{-0.25}$	$0.34^{+0.45}_{-0.34}$	—
B	bin 6	$0.018^{+0.029}_{-0.012}$	$0.00033^{+0.00011}_{-0.00029}$	$0.019^{+0.063}_{-0.012}$	98%
B	bin 7	$0.027^{+0.264}_{-0.018}$	$0.46^{+0.38}_{-0.22}$	$0.027^{+0.111}_{-0.018}$	1600%
B	bin 8	$1.36^{+0.88}_{-0.7}$	$1.63^{+0.77}_{-0.43}$	$0.84^{+0.53}_{-0.39}$	—
B	bin 9	$0.0133^{+0.0034}_{-0.009}$	$0.049^{+0.305}_{-0.049}$	$0.015^{+0.121}_{-0.011}$	—
B	bin 10	$0.14^{+0.37}_{-0.14}$	$0.15^{+0.48}_{-0.13}$	$0.023^{+0.171}_{-0.022}$	—

Table 6.30: Background expectations in the signal mass window [492,504] MeV/c^2, using the base model and two alternative models (sum of a power-law and linear regression and base model with reduced lower sideband), for the TOS trigger category. In case the asymmetric uncertainties were not available, the symmetric ones are shown.

sample	bin	[480,492] MeV/c^2		[504,516] MeV/c^2		absolute	
		fitted	actual	fitted	actual	systematics	
A	bin 1	$11.9^{+3.1}_{-5.0}$	9	$1.41^{+0.94}_{-0.63}$	3	—	
A	bin 2	$3.0^{+5.3}_{-1.3}$	5	$0.2^{+0.3}_{-0.13}$	2	—	+3.71
A	bin 3	$2.4^{+2.6}_{-1.1}$	10	$1.18^{+1.04}_{-0.6}$	1	−1.26	+7.74
A	bin 4	$3.9^{+3.4}_{-1.4}$	10	$1.3^{+1.15}_{-0.74}$	0	—	
A	bin 5	$2.3^{+2.2}_{-1.6}$	5	$0.82^{+0.74}_{-0.41}$	0	—	
A	bin 6	$2.4^{+2.5}_{-1.2}$	0	$0.22^{+0.32}_{-0.22}$	0	—	
A	bin 7	$0.26^{+0.83}_{-0.17}$	0	$0.00017^{+0.00112}_{-0.00014}$	0	—	
A	bin 8	$2.4^{+2.8}_{-1.6}$	0	$0.000373^{+8.4e-05}_{-0.00029}$	1	−0.31	+0.69
A	bin 9	$0.038^{+0.592}_{-0.03}$	0	$5.3e-06^{+8.5e-06}_{-3.5e-06}$	0	—	
A	bin 10	$0.109^{+0.213}_{-0.075}$	0	$8.8e-05^{+0.000188}_{-5.2e-05}$	0	—	
B	bin 1	$5.2^{+2.6}_{-1.8}$	4	$3.43^{+0.51}_{-0.43}$	0	−0.6	+3.4
B	bin 2	$9.4^{+4.8}_{-3.7}$	7	$1.6307^{+1.1919}_{-0.0034}$	2	—	
B	bin 3	$5.8^{+3.5}_{-3.1}$	3	$0.73^{+0.75}_{-0.58}$	1	—	
B	bin 4	$7.4^{+4.9}_{-3.4}$	3	$0.0116^{+0.0016}_{-0.0116}$	2	—	+1.92
B	bin 5	$1.18^{+2.04}_{-0.67}$	3	$0.17^{+0.24}_{-0.11}$	0	—	
B	bin 6	$1.3^{+1.39}_{-0.66}$	6	$1.45^{+0.85}_{-0.95}$	0	−1.07	+4.93
B	bin 7	$0.76^{+2.21}_{-0.65}$	1	$0.037^{+0.163}_{-0.037}$	0	—	
B	bin 8	$0.56^{+0.82}_{-0.32}$	1	$0.03^{+0.272}_{-0.03}$	0	—	
B	bin 9	$0.118^{+0.295}_{-0.067}$	1	$0.28^{+0.35}_{-0.24}$	0	—	
B	bin 10	$1.6^{+2.51}_{-0.91}$	1	$0.37^{+0.44}_{-0.28}$	0	—	

Table 6.31: Number of background events expected in the mass window [480,492] MeV/c^2 ([504,516] MeV/c^2) from a fit of the base model on the region [470,480]∪[504,600] MeV/c^2 ([470,492]∪[516,600] MeV/c^2), compared to the actual number of events observed in this window, for TIS triggers. In case the asymmetric uncertainties were not available, the symmetric ones are shown. In case there is a notable discrepancy between the value from the fit and the actual number of event, a systematical uncertainty is computed.

sample	bin	[480,492] MeV/c^2		[504,516] MeV/c^2		absolute
		fitted	actual	fitted	actual	systematics
A	bin 1	$10.9^{+4.9}_{-3.4}$	10	$1.38^{+1.28}_{-0.71}$	0	—
A	bin 2	$4.4^{+2.6}_{-1.7}$	9	$0.22^{+0.45}_{-0.1}$	1	—
A	bin 3	$1.91^{+1.98}_{-0.89}$	3	$0.31^{+0.62}_{-0.31}$	0	—
A	bin 4	$3.2^{+2.4}_{-1.7}$	3	$0.11^{+0.24}_{-0.11}$	0	—
A	bin 5	$11.3^{+2.4}_{-2.2}$	2	$0.38^{+0.51}_{-0.38}$	0	-0.35 $+1.65$
A	bin 6	$6.2^{+3.9}_{-6.2}$	0	$0.000306^{+7.8e-05}_{-0.000228}$	1	-0.28 $+0.72$
A	bin 7	$3.7^{+2.7}_{-1.8}$	1	$0.028^{+0.319}_{-0.021}$	0	—
A	bin 8	$1.5^{+2.7}_{-1.5}$	3	$0.48^{+0.56}_{-0.34}$	0	—
A	bin 9	$0.45^{+0.54}_{-0.26}$	0	$4.5e-05^{+3.1e-05}_{-3.3e-05}$	0	—
A	bin 10	$8.9^{+5.1}_{-3.9}$	2	$0.39^{+0.48}_{-0.26}$	0	-0.68 $+1.38$
B	bin 1	$8.2^{+4.6}_{-2.6}$	9	$1.21^{+1.39}_{-0.65}$	0	—
B	bin 2	$7.0^{+4.5}_{-3.3}$	5	$1.08^{+0.86}_{-0.62}$	1	—
B	bin 3	$5.3^{+3.9}_{-5.3}$	3	$0.38^{+0.68}_{-0.38}$	0	—
B	bin 4	$4.0^{+4.2}_{-1.4}$	6	$0.45^{+0.65}_{-0.45}$	0	—
B	bin 5	$2.02^{+0.26}_{-0.79}$	4	$0.126^{+0.472}_{-0.084}$	0	—
B	bin 6	$1.81^{+2.34}_{-0.79}$	0	$0.000495^{+6.1e-05}_{-0.000337}$	0	—
B	bin 7	$1.31^{+0.2}_{-1.31}$	1	$0.00105^{+0.00024}_{-0.00076}$	0	—
B	bin 8	$7.1^{+3.2}_{-2.8}$	3	$0.24^{+0.46}_{-0.2}$	1	—
B	bin 9	$0.71^{+1.27}_{-0.38}$	1	$0.00044^{+0.00022}_{-0.00032}$	0	—
B	bin 10	$0.26^{+1.22}_{-0.18}$	1	$0.048^{+0.277}_{0.048}$	0	—

Table 6.32: Number of background events expected in the mass window [480,492] MeV/c^2 ([504,516] MeV/c^2) from a fit of the base model on the region [470,480]∪[504,600] MeV/c^2 ([470,492]∪[516,600] MeV/c^2), compared to the actual number of events observed in this window, for TOS triggers. In case the asymmetric uncertainties were not available, the symmetric ones are shown. In case there is a notable discrepancy between the value from the fit and the actual number of event, a systematical uncertainty is computed.

6.9 Results

Once all the normalisation factors and expected backgrounds in each BDT bin have been calculated, the "unblinding" has been performed and the observed number of events obtained. The result is that no significant signal excess has been observed, so upper limits on the $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ have been set. Table 6.33 shows the expected and observed number of events in the 10 BDT bins of sample A and B of TIS and TOS categories. The expected numbers include all the systematics errors explained in the previous section.

BDT bin		Sample A		Sample B	
		TIS	TOS	TIS	TOS
1	Exp. bkg	$3.2^{+1.4}_{-3.2}$	$2.1^{+2.5}_{-2.1}$	$0.60^{+3.41}_{-0.60}$	$1.66^{+1.10}_{-0.78}$
	Observed	0	1	2	4
2	Exp. bkg	$1.29^{+3.89}_{-0.66}$	$0.86^{+0.73}_{-0.39}$	$2.75^{+1.36}_{-0.64}$	$1.51^{+1.14}_{-0.75}$
	Observed	3	2	3	2
3	Exp. bkg	$2.3^{+7.8}_{-1.5}$	$0.23^{+0.40}_{-0.23}$	$1.15^{+0.99}_{-0.64}$	$0.39^{+0.80}_{-0.39}$
	Observed	1	0	3	1
4	Exp. bkg	$1.44^{+1.24}_{-0.75}$	$0.23^{+0.50}_{-0.23}$	$1.08^{+2.12}_{-0.58}$	$0.46^{+0.55}_{-0.21}$
	Observed	0	0	1	1
5	Exp. bkg	$1.14^{+0.73}_{-0.54}$	$0.35^{+1.73}_{-0.35}$	$0.35^{+0.44}_{-0.35}$	$0.30^{+0.45}_{-0.20}$
	Observed	1	0	0	0
6	Exp. bkg	$0.25^{+0.41}_{-0.25}$	$0.28^{+0.85}_{-0.28}$	$1.1^{+5.0}_{-1.1}$	$0.018^{+0.034}_{-0.018}$
	Observed	0	0	0	2
7	Exp. bkg	$0.00139^{+0.00028}_{-0.00111}$	$0.21^{+0.36}_{-0.14}$	$0.102^{+0.342}_{-0.096}$	$0.027^{+0.506}_{-0.027}$
	Observed	0	0	0	0
8	Exp. bkg	$0.31^{+0.84}_{-0.31}$	$0.59^{+0.80}_{-0.59}$	$0.081^{+0.385}_{-0.081}$	$1.36^{+0.88}_{-0.70}$
	Observed	1	0	0	0
9	Exp. bkg	$0.00028^{+0.00645}_{-0.00018}$	$0.0027^{+0.0026}_{-0.0027}$	$0.23^{+0.37}_{-0.23}$	$0.0133^{+0.0034}_{-0.0090}$
	Observed	0	0	0	0
10	Exp. bkg	$0.000193^{+0.000046}_{-0.000173}$	$0.68^{+1.54}_{-0.68}$	$0.37^{+0.63}_{-0.37}$	$0.14^{+0.37}_{-0.14}$
	Observed	0	1	0	0

Table 6.33: Expected background and observed $K^0_S \to \mu^+\mu^-$ candidates in per BDT bin in each of the four samples: TIS–TOS in samples A and B.

In order to obtain limits on the $\mathcal{B}(K^0_S \to \mu^+\mu^-)$, the CL_s method has been used. This was also the choice in the $B^0_{(s)} \to \mu^+\mu^-$ analysis, and a brief explanation on it was given in section 5.7.2.1. The input to the method are the expected and observed backgrounds per BDT bin shown in table 6.33 and the normalisation factors computed in section 6.7. It is also important to recall that the outputs were the CL_s (compatibility of the observed distribution with the signal hypothesis) and the CL_b (compatibility with the background-only hypothesis). CL_s is used to set the upper limits on the branching fractions while $1 - CL_b$ gives a $p - value$ needed for an evidence or observation.

As a reminder, the systematic errors considered for the limit (with the corresponding correlations) are:

- Background expectations per bin. These are taken from table 6.33. No correlation is considered among the different bins.

- Ratios of reconstruction, selection, and muon identification efficiencies. Those are taken from table 6.26 and 6.28. No correlation is considered among the different bins.

- $\mathcal{B}(K^0_S \to \pi^+\pi^-)$, $(69.20 \pm 0.20)\%$. This systematic source affects simultaneously the signal expectations of the 40 bins of the analysis.
- Validation of the ratio of TIS efficiencies using simulation, computed in section 6.7.5. This error affects coherently the signal expectations of the 20 bins of the TIS analysis.
- Effective prescale factor of the MB sample, obtained in section 6.7.6. This error affects coherently the signal expectations of the 20 bins of the TOS analysis.

Following all this, figure 6.38 shows the expected and observed CL_s vs. $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ curves for the TIS and xTOS categories, as well as for the final combined result. As already said, the observed distribution of events is compatible with background expectations, giving a p-value of $\sim 27\%$. The upper limit on $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ found is 11.2 (9.0)$\times 10^{-9}$ at 95 (90)% CL, slightly above the expected 11.2 (8.7) $\times 10^{-9}$ and a factor of 35 better than previous world best limit in [195]. Table 6.34 summarises the limits in the TIS, xTOS, and combined analysis.

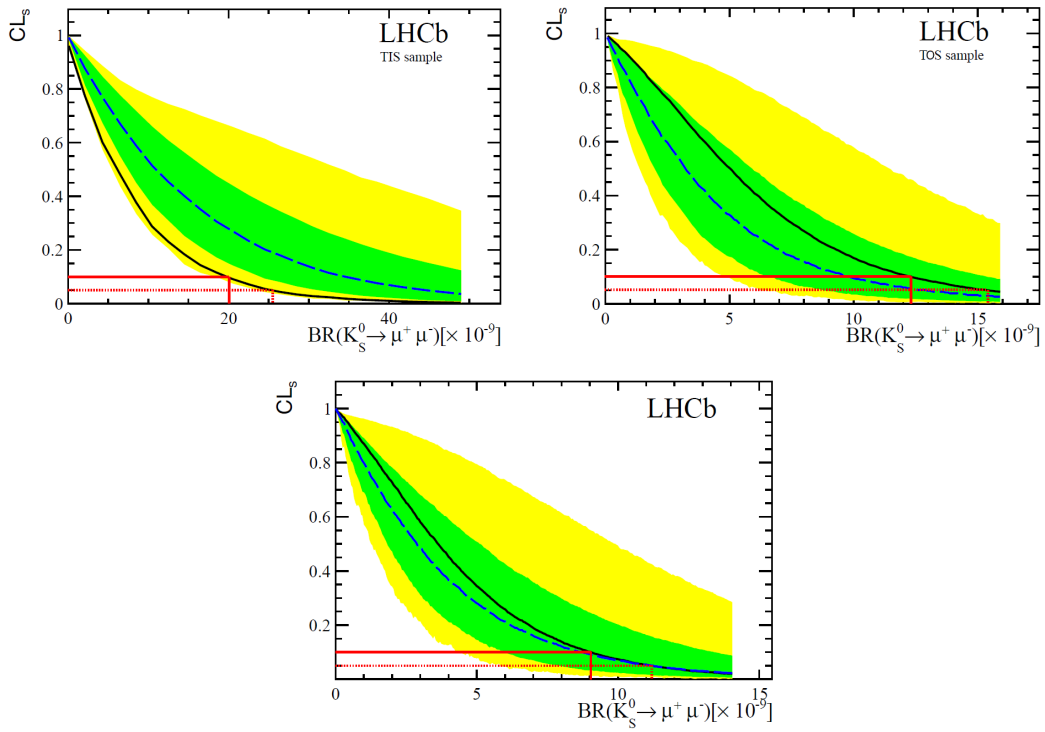

Figure 6.38: CL_s vs. $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ curves for TIS (top left), xTOS (top right) and combined (bottom) analyses. The solid line corresponds to the observed CL_s. The dashed line corresponds to the median of the CL_s for an ensemble of background-alone experiments. In each plot, two bands are shown. The yellow band covers 68% of the CL_s curves obtained in the background alone experiment, while the green band covers 95%.

Quantity	TIS	xTOS	Combined
Expected 95 (90)% upper limit	44.4 (34.5)	12.7 (9.8)	11.2 (8.7)
Observed 95 (90)% upper limit	**25.5 (20.1)**	**15.4 (12.3)**	**11.2 (9.0)**
p-value	0.95	0.2	0.27

Table 6.34: Upper limits on $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ in units of 10^{-9} for the TIS, xTOS, and combined samples. The last entry in the table is the p-value of the background-only hypothesis.

As already explained, the reported search for $K^0_S \to \mu^+\mu^-$ has not yielded any significant excess above the background. However, in the 6^{th} BDT bin of sample B TOS, 2 events have been observed for a expectancy of $0.018^{+0.034}_{-0.018}$ background (see table 6.33). This is not significant when the the rest of the bins are taken into account, but still a small study of these two observed candidates has been performed. In one of the two $K^0_S \to \mu^+\mu^-$ candidates, the two muons are clones sharing the same muon hits, so background. However, the other candidate has been found to be very signal like, with two well separated muons, mass far from the K^0_S one in the $\pi^+\pi^-$ mass hypothesis and close to it in the $\mu^+\mu^-$ one, and lifetime also similar to the K^0_S one. Figure 6.39 shows an event display of the candidate, where the muons of the final state can be easily distinguished.

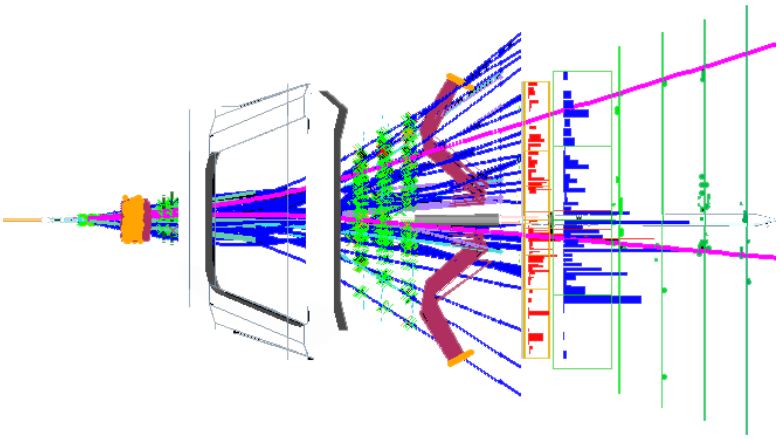

Figure 6.39: Event display of a potential candidate for $K^0_S \to \mu^+\mu^-$, built with Panoramix. Both muons of the final state can be easily distinguished (magenta lines).

7 Conclusions

This thesis has presented the searches for $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$ using the 1 fb^{-1} of data collected by LHCb during 2011 from the LHC proton–proton collisions at a center of mass energy of $\sqrt{s} = 7$ TeV.

It has been seen that the LHCb experiment showed an excellent performance during the year 2011, this being needed for searches such as $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$. This excellent performance includes trigger, vertexing, tracking and PID, all of them crucial for the experimental results reported.

The rare decays searches by LHCb have been overviewed, and the most important results on it showed. In particular, the angular analysis in $B^0 \to K^{*0}\mu^+\mu^-$ yields some of the most interesting results in this area. Indeed, even if these results are compatible with SM, they produce important constraints in the NP parameter space. A promising result is also produced by the isospin asymmetry searches with $B \to K\mu^+\mu^-$ modes, although more theory work is required here.

The $B^0_{(s)} \to \mu^+\mu^-$ decays have been showed to be one of the golden modes for NP searches in the flavour sector. This is due to the fact that their $\mathcal{B}$s are sensitive to new particles entering in the Feynman diagrams loops, therefore allowing for an indirect search of NP reaching energies larger than in the direct searches. The SM predictions on the $\mathcal{B}$s were $\mathcal{B}(B^0 \to \mu^+\mu^-)|_{SM} = (1.0 \pm 0.1) \times 10^{-10}$ and $\mathcal{B}(B^0_s \to \mu^+\mu^-)|_{SM} = (3.2 \pm 0.2) \times 10^{-9}$.

The LHCb search for $B^0_{(s)} \to \mu^+\mu^-$ decays is based on the use of a loose selection which removes most of the background and in the classification of the surviving candidates in a 2D space which uses a BDT and the dimuon invariant mass as variables. The compatibility of the number of observed candidates in each bin of this space with the the expected background and signal hypotheses is obtained with the CL$_s$ method. In order to convert a signal hypothesis into a $\mathcal{B}$ a normalisation to a channel with known $\mathcal{B}$ is used. In the case of $B^0_{(s)} \to \mu^+\mu^-$, the normalisation is done to $B^+ \to J/\psi K^+$, $B^0_s \to J/\psi\,\phi$ and $B^0_{(s)} \to h^+h'^-$.

The search using the 2011 LHCb dataset has not produced any significant signal excess, so upper limits to the $\mathcal{B}$ of both $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ have been set:

$$\boxed{\mathcal{B}(B^0 \to \mu^+\mu^-) < 1.0 \times 10^{-9} \text{ at } 95\% \text{ CL}}$$

$$\boxed{\mathcal{B}(B^0_s \to \mu^+\mu^-) < 4.5 \times 10^{-9} \text{ at } 95\% \text{ CL}}$$

Both these results are world best, and in the case of $B_s^0 \to \mu^+\mu^-$ discard a possible $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ much larger than the SM prediction.

The $K_S^0 \to \mu^+\mu^-$ decay has been showed to be dominated by LD diagrams in SM, but also to have a SD contribution which could be enhanced by NP models, promoting the channel $\mathcal{B}$ almost an order of magnitude above the SM prediction, which is $\mathcal{B}(K_S^0 \to \mu^+\mu^-)|_{SM} = (5.1 \pm 1.5) \times 10^{-12}$.

The search for $K_S^0 \to \mu^+\mu^-$ at LHCb has been showed to be challenging because of the long K_S^0 distance of flight (which makes many of these particles decay outside the VELO) and also because of the low p_T of its daughters. This produces low trigger and reconstruction efficiencies, and strong biases that have to be dealt with. As an example, the analysis had to be divided depending on if the K_S^0 was responsible for the trigger or some other particle in the underlying event was. Apart from these and other details, the analysis strategy has been showed to be similar to the one in $B_{(s)}^0 \to \mu^+\mu^-$, using the CL_s method, a BDT discriminant and normalising to a channel with known $\mathcal{B}$. In this case, the channel chosen has been $K_S^0 \to \pi^+\pi^-$.

No significant signal excess has been found in the search for $K_S^0 \to \mu^+\mu^-$ using the 1 fb^{-1} collected by LHCb in 2011, so upper limits in the $\mathcal{B}(K_S^0 \to \mu^+\mu^-)$ have been set:

$$\boxed{\mathcal{B}(K_S^0 \to \mu^+\mu^-) < 11.2 \times 10^{-9} \text{ at } 95\% \text{ CL}}$$

This upper limits turns out to be a factor of 35 better than the previous world best.

8 Summary

The Standard Model (SM) of Particle Physics turns out to be the most reliable theory to explain the structure and relations of all the fundamental particles. These include the nucleons and electrons that compose the atoms, photons and a whole set of new particles discovered during 20^{th} century and beginning of 21^{st}.

In spite of its success, there are still certain experimental measurements which the SM fails to completely explain. Because of this, several New Physics (NP) models have been proposed to solve these issues.

Experiments such as the LHC, at CERN, have been created to test the SM and NP models predictions. LHC collides protons head–to–head at a energy of $\sqrt{s} = 7$ TeV, and several detectors register the result of these collisions. Among these detectors is LHCb, designed to study the CP–violation and rare decays such as $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_{\rm S} \to \mu^+\mu^-$.

8.1 Theory aspects of $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_{\rm S} \to \mu^+\mu^-$

$B^0_{(s)} \to \mu^+\mu^-$ and $K^0_{\rm S} \to \mu^+\mu^-$ are very rare decays, with a branching ratio, $\mathcal{B}$, in the SM suppressed down to the $10^{-9} - 10^{-12}$ level. These models are, however, sensitive to NP, so that its $\mathcal{B}$ could be enhanced (or even more suppressed) with respect to the SM predictions. Therefore, the measurement of a $\mathcal{B}$ significantly different to the SM one would be a sign of physics beyond the SM.

$B^0_{(s)} \to \mu^+\mu^-$ and $K^0_{\rm S} \to \mu^+\mu^-$ modes are flavour changing neutral currents, so forbidden at tree level within the SM framework. The diagrams through which these decays are produced include *loops* and *boxes*, in which non–SM particles can enter and change, as explained, the actual value of the branching fractions with respect to the SM prediction. Because of this, these decays are suitable for an indirect search for NP, with the advantage of reaching higher energy scales than in direct searches.

$B^0_{(s)} \to \mu^+\mu^-$ decays are among the golden channels for the search of NP in the flavour sector. Their $\mathcal{B}$ have been widely studied, and accurate predictions in the SM have been set to $\mathcal{B}(B^0 \to \mu^+\mu^-)|_{SM} = (1.1 \pm 0.1) \times 10^{-10}$ and $\mathcal{B}(B^0_s \to \mu^+\mu^-)|_{SM} = (3.2 \pm 0.3) \times 10^{-9}$. As mentioned, both these branching ratios are very sensitive to the presence of NP, and there is plenty of literature on the theory to see how these $\mathcal{B}$s can be affected depending on the NP phase space. Table 8.1 summarises some of these results in the $B^0_s \to \mu^+\mu^-$ case, whose $\mathcal{B}$ is higher in the SM (so easier to be tested by the experiments).

$K^0_{\rm S} \to \mu^+\mu^-$ decay has been proposed as novel way to test the SM using the potential

Scenario	Would point to
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \gg$SM	Big enhancements from NP in scalar sector, SUSY with high $\tan\beta$
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \neq$SM	SUSY, EDs, LHT, TC2
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \sim$SM	rule out regions of parameter space that predict sizeable departures from SM
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \ll$SM	NP in scalar sector, but MSSM ruled out or quite unlikely. NMSSM good candidate

Table 8.1: Implications for NP of different measured $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$

of the LHCb experiment. At SM, $K_S^0 \to \mu^+\mu^-$ decay is basically dominated by long–distance physics, with a much smaller short–distance contribution. This short–distance part is, nevertheless, sensitive to NP, and could become even dominant depending on the NP phase space, promoting the channel $\mathcal{B}$ almost an order of magnitude above the SM prediction, which is $\mathcal{B}(K_S^0 \to \mu^+\mu^-)|_{SM} = (5.1 \pm 1.5) \times 10^{-12}$. $K_S^0 \to \mu^+\mu^-$ has been suggested as as a possible way to look for new light scalars and also to set model-independent bounds on the CP-violating phase of the $s \to dl^+l^-$ amplitude.

8.2 The LHCb experiment at the LHC

LHCb is one of the four big detectors of the LHC experiment at CERN. LHCb is a specific purpose detector: study of CP violation and rare decays, mainly from decays of particles with bottom quark.

LHCb has been designed to work at an instant luminosity smaller than that of other LHC experiments. The reason to do so is easing the correct identification of primary vertex (PV) and secondary vertex (SV), as this is essential for the LHCb physics goals.

8.2.1 LHCb detector

LHCb is a single-arm spectrometer with a forward angular coverage from approximately 10 mrad to 300 (250) mrad in the bending (non-bending) plane (see figure 8.1). Main elements of LHCb are:

- a spectrometer magnet, a warm dipole magnet providing an integrated field of 4 T·m,
- a vertex locator (VELO) system,
- a tracking system, made of the Tracker Turiciensis (a silicon microstrip detector, TT) in front of the spectrometer magnet, and three tracking stations behind the magnet, made of silicon microstrips in the inner parts (IT) and of gas straws for the outer parts (OT),
- two Ring Imaging Cherenkov counters (RICH1 and RICH2), to achieve excellent p–K separation in the momentum range from 2 to 100 GeV/c,
- a calorimeter system composed of a Scintillator Pad Detector and Preshower (SPD/PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL),
- a muon detection system composed of MWPC.

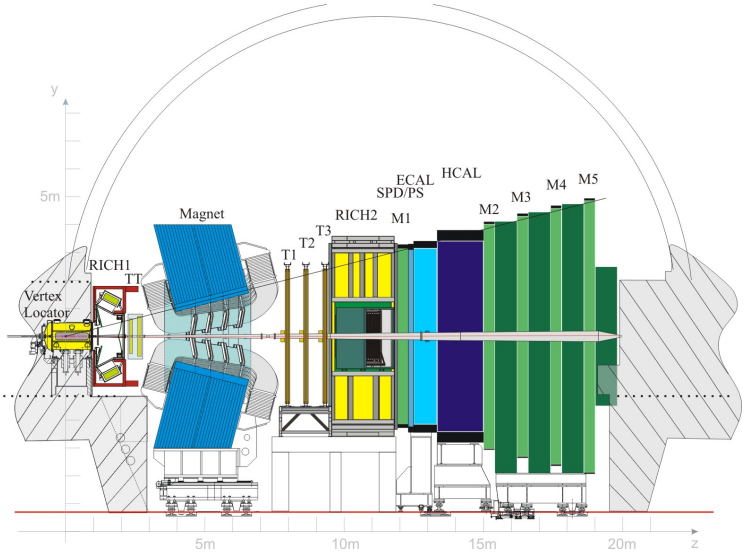

Figure 8.1: Side view of LHCb

8.2.2 LHCb analysis workflow

Data taken by LHCb from LHC proton–proton collisions, at a rate of several million events per second, need to be selected, prepared and distributed in the most possible efficient way for its subsequent analysis. This process is summarised next:

1. The amount of data generated by LHC collisions is too high to be directly stored. In this way, it becomes necessary an initial and fast selection which allows to discard most of the events that are not interesting for the physics analysis. This initial selection is called trigger, and it is common to all high energy physics experiments. In particular, during 2011 LHCb trigger reduced the rate from ∼15 MHz to ∼3 kHz.

2. Data selected by the trigger are transformed by different mathematical algorithms in an ensemble of tracks and vertices. The reconstruction of tracks and vertices is mainly done with the tracking detectors (including VELO). The information from the Particle Identification (PID) detectors is then added to identify the nature of the tracks, distinguishing, for instance, muons from other particles.

3. Once all the triggered events have been reconstructed, it becomes necessary to separate them according to their physics content. This is done by selecting the different decays using their particular features. For instance $B^0_{(s)} \to \mu^+\mu^-$ or $K^0_S \to \mu^+\mu^-$, with same final state, have different mass and lifetime of the mother, expected separation of the daughters with respect to the PV,...

4. The triggered, reconstructed and selected dataset has to be then distributed to a series of computing centers spread worldwide. With this, the access to the data becomes easier and its analysis can be distributed. This is all achieved through the Grid system.

5. The final tool needed to produce physics results is the use of simulated Monte Carlo (MC) data. This is crucial in order to understand several effects and biases created by the group of processes just described. The production of MC data as similar as possible to the real one is one of the main efforts of all the high energy physics experiments.

The analyses of this thesis are based on the data taken by LHCb from the LHC proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 7$ TeV during 2011. The luminosity in this period was kept in the $3 - 3.5 \times 10^{32}$ cm^{-2} s^{-1} range. The integrated luminosity corresponds to ~ 1.02 fb^{-1}.

8.3 Search for rare decays at LHCb

Some of the most remarkable LHCb results so far come in the rare decays area. The rare decays analyses include several common experimental features, what makes easy to share tools and experience on solving issues among all of them. The rare decays searches at LHCb are typically divided in electroweak penguin transitions $b \to d\mu^+\mu^-$ and $b \to s\mu^+\mu^-$, radiative decays and very rare decays.

Electroweak penguin transitions in decays such as $b \to d\mu^+\mu^-$ and $b \to s\mu^+\mu^-$ produce a series of observables very sensitive to NP. In the particular case of $B^0 \to K^{*0}\mu^+\mu^-$ these observables are accessible through the use of the angular distribution of the decay products, and are dependent on the invariant mass of the dimuon pair. This dependence is well predicted in the SM and differs in other models. LHCb has measured these observables with the 2011 dataset, and results obtained are world best and compatible with the SM expectancy. Another interesting observable of the same kind is the isospin asymmetry measurements in $B \to K\mu^+\mu^-$ and $B \to K^{*0}\mu^+\mu^-$ modes. Using 2011 dataset, LHCb has measured this asymmetry to be compatible with 0 in the $B \to K^{*0}\mu^+\mu^-$ case (in agreement with the SM prediction), but significantly below for $B \to K\mu^+\mu^-$. Even if there is no a SM prediction yet on the second mode, this is expected to be also 0. As a final example of the work in electroweak penguin transitions, LHCb has discovered the rarest B meson decay seen up to now, $B^+ \to \pi^+\mu^+\mu^-$, with a $\mathcal{B}$ at the level of $\sim 2.4 \times 10^{-8}$.

The radiative b decays could also show hints of physics beyond the SM, since NP can introduce sizeable effects on the dynamics of the radiative transitions. An example of observable sensitive to these NP effects is the direct CP asymmetry in $b \to s\gamma$ using the $B^0 \to K^{*0}\gamma$ channel. Thanks to the 1 fb^{-1} collected by LHCb during 2011, this CP asymmetry has been measured to be consistent with the SM prediction, very close to 0.

The "very rare decays" tag refers to models extremely suppressed in the SM, with potential enhancements (or even further suppressions) in NP models, so that a direct observation or upper limits in the $\mathcal{B}$ can set bounds in the NP phase space. $B^0_{(s)} \to \mu^+\mu^-$ and $K^0_S \to \mu^+\mu^-$ are examples of very rare decays. Apart from those, LHCb has searched for more very rare decays using the 2011 dataset. Examples are the lepton flavour violating decays $\tau^- \to \mu^+\mu^-\mu^-$ and $B^- \to X^+\mu^-\mu^-$, and also other channels such as $B^0_{(s)} \to \mu^+\mu^-\mu^+\mu^-$ and $D^0 \to \mu^+\mu^-$. LHCb has not found any significant signal excess in these modes, but has set world best (or close to world best) upper limits in the branching fractions of all of them.

8.4 $B^0_{(s)} \to \mu^+\mu^-$

Due to their interest, the $B^0_{(s)} \to \mu^+\mu^-$ decays have been searched by many experiments before. The best upper limits at the beginning of 2012 were those given by LHCb, using around 30% of the 2011 dataset: $\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.6 \times 10^{-9}$ and $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 1.6 \times 10^{-8}$, both at 95 % CL.

For the search for these channels at LHCb a very efficient selection has been designed to remove most of the background. The number of observed events is then compared to the number of expected signal and background events in bins of two independent variables, the invariant mass and the output of a multi-variate discriminant, a Boosted Decision Tree (BDT). The probability for a signal or background event to have a given value of BDT is extracted from data using $B^0_{(s)} \to h^+h'^-$ candidates for signal and from $B^0_{(s)} \to \mu^+\mu^-$ sidebands for combinatorial background. The invariant mass line shape of the signals is described by a Crystal Ball whose parameters are also extracted from data. For the background, the line shape is extracted again from the $B^0_{(s)} \to \mu^+\mu^-$ sidebands. The number of expected signal events, for a given branching fraction hypothesis, is obtained by normalising to channels with known branching fractions: $B^+ \to J/\psi K^+$, $B^0_s \to J/\psi\,\phi$ and $B^0 \to K^+\pi^-$.

Several specific backgrounds have been tested to be negligible for the analysis. The only exception is $B^0_{(s)} \to h^+h'^-$ in which both hadrons are misidentified as muons. Even if the

contribution of this background is small thanks to the PID cuts, this has also been parameterised in bins of mass and BDT.

Once the observed and expected number of signal and background candidates are available in each bin of the BDT–invariant mass space, their compatibility is calculated using the CL_s method. With this, as the expected number of signal is proportional to the branching ratio hypothesis, upper limits on it can be set at a given confidence level.

It is also important to say this analysis is blinded, i.e., the signal mass region is not looked at until the completion of the analysis.

8.4.1 Selection and BDT

The $B^0_{(s)} \to \mu^+\mu^-$ selection is designed to be similar in the signal and control channels, in order to reduce the systematic errors when performing the normalisation and calibration. In particular, the only difference between $B^0_{(s)} \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h'^-$ selection are the PID cuts applied to the muons in the first case. It is important to say that an initial BDT using simple variables, called BDTS, is also introduced at selection level. The BDTS is designed to behave similarly in normalisation and signal channels.

The BDT used to classify the events surviving the selection has been studied in detail, in order to search for the best parameterisation and choice of variables producing the best result. In this way, the final choice included 9 different variables, gathering properties of the muons and the B meson. The BDT training is done using MC $B^0 \to \mu^+\mu^-$ for signal and MC $b\bar{b} \to \mu\mu X$ for background.

A study has also been performed to find the BDT (and mass) binning producing the best sensitivity.

8.4.2 Specific backgrounds

The most relevant background for this analysis is the combinatorial one, composed mainly by $b\bar{b} \to \mu\mu X$, i.e., pairs of b quarks which decay independently to a muon.

The most important specific background is the $B^0_{(s)} \to h^+h'^-$ in which both hadrons are misidentified as muons. The contribution of this background has been obtained by measuring the misidentification rate from kaons and pions in $D^0 \to K^+\pi^-$ as a function of p, p_T and applying the obtained mapping to the $B^0_{(s)} \to h^+h'^-$ phase space, obtained from MC. The total expected amount of background is obtained from the total $B^0_{(s)} \to h^+h'^-$ measured in the 2011 data sample. Its BDT and mass distributions are obtained as well from MC.

Other specific backgrounds such as $B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$, $B^+ \to \pi^+\mu^+\mu^-$ and $B_s \to \mu^+\mu^-\gamma$ have been estimated to be negligible for the current search.

8.4.3 Calibration of BDT and invariant mass

The BDT PDF for the signals is extracted using events from the inclusive $B^0_{(s)} \to h^+h'^-$ sample in which the B meson was not responsible for the trigger. This BDT PDF is extracted using different fitting functions, with the measured differenced included as a systematic error. An example of these fits for the last two BDT bins can be found in figure 8.2. The BDT shape produced is then corrected to take into account the effect of the trigger and muon identification (muonID) cuts.

The mass PDFs for the signals are parameterised as two Crystal Ball functions. The central values are extracted from the exclusive $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$ samples, while the resolutions are measured by interpolating at the B^0 and B_s^0 mass point the resolutions measured with the J/ψ, $\Psi(2S)$, $\Upsilon(1S, 2S, 3S)$ dimuon resonances

The BDT and invariant mass PDFs for the combinatorial background are extracted from of a fit of the dimuon mass sidebands in each BDT bin, using an exponential PDF and other models as a cross-check.

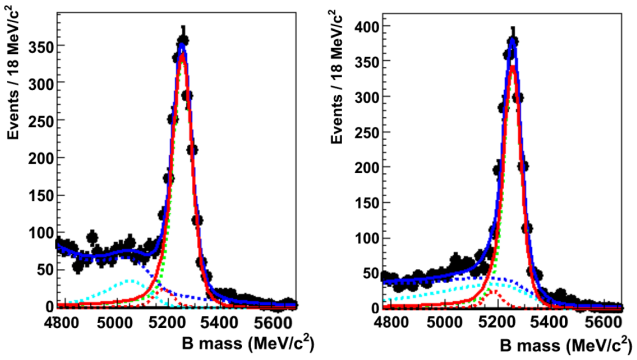

Figure 8.2: Fits to the $B^0_{(s)} \to h^+h'^-$ mass spectra to obtain the BDT calibration. This example shows the last 2 BDT bins.

8.4.4 Normalisation

The normalisation to a channel with known branching ratio permits converting a number of candidates in a $\mathcal{B}$ without the knowledge of the absolute luminosity and $b\bar{b}$ cross section. The channels chosen for normalisation at LHCb are: $B^+ \to J/\psi K^+$, $B^0_s \to J/\psi\,\phi$ and $B^0 \to K^+\pi^-$. Therefore, the normalisation requires measuring the ratios of reconstruction, selection and trigger efficiencies of the signal and the mentioned calibration channels, as well as the ratio of production fractions (depending on the calibration channel) and the number of calibration candidates:

- Production fractions: the ratio $f_{B^0}/f_{B^0_s}$ measured by LHCb is used, whenever the normalisation implies two modes with different mothers.
- Reconstruction: for the reconstruction MC simulation is used, having the result validated in data.
- Selection: for the geometrical cuts, once again MC simulation is used, but with the parameters "smeared" to match the resolutions in data. For the PID, the ratios are extracted from data combined with MC.
- Trigger: the trigger efficiency for the normalisation channels with muons is directly obtained from data. For the signal, weighting the efficiency measured in data to the p and IP spectra of $B^0 \to \mu^+\mu^-$ in MC. In the case of $B^0 \to K^+\pi^-$, the normalisation is done to the cases in which the B^0 meson is not responsible for the trigger decision, with the efficiency for that measured using $B^+ \to J/\psi K^+$ decays.
- Number of candidates: the number of candidates for the normalisation channels is obtained fitting the invariant mass spectra in each case.

8.4.5 Results

The search for $B^0_{(s)} \to \mu^+\mu^-$ resulted in no significant signal excess. Figure 8.3 shows the invariant mass distribution of the observed candidates and the expected background and SM signal contributions in the BDT range above 0.5.

Since no signal excess has been found, upper limits have been set in the branching ratios of both channels. As explained, the CL_s method is used in order to obtain these limits. Both results, showed in table 8.2, are world best, and produce important bounds to the NP phase space.

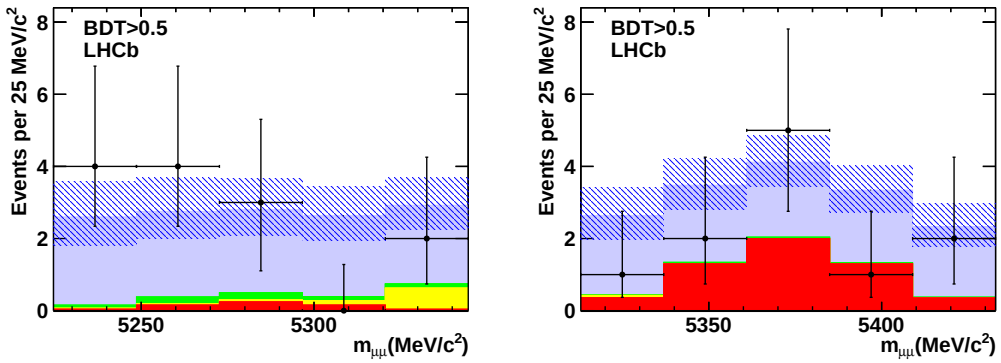

Figure 8.3: $B^0 \to \mu^+\mu^-$ (left) and $B^0_s \to \mu^+\mu^-$ (right) distribution of selected dimuon events in the invariant mass plane for events with BDT>0.5. Data and different background contributions are shown. The effect of a potential SM signal can be also seen. The shaded area reflects the error in the sum of all the contributions.

Channel	95% CL limit	90 % CL limit
$B^0 \to \mu^+\mu^-$	1.0×10^{-9}	0.8×10^{-9}
$B^0_s \to \mu^+\mu^-$	4.5×10^{-9}	3.8×10^{-9}

Table 8.2: Upper limits on the $\mathcal{B}$ of $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ decays using 2011 LHCb dataset.

8.5 $K^0_S \to \mu^+\mu^-$

The experimental upper limit in the $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ is 3.2×10^{-7} at 90 % CL, obtained almost 40 years ago.

The search for $K^0_S \to \mu^+\mu^-$ is challenging due to the fact that LHCb is optimised for B mesons, which show different features to K^0_S. As an example, the K^0_S mass is around 10 times smaller than that of B^0, while it flies a factor of more than 50 more. This implies very different reconstruction and trigger performances.

The $K^0_S \to \mu^+\mu^-$ analysis at LHCb also starts with an efficient selection to get rid of most of the background. However, the strong biases caused by the low trigger efficiency force the split of the analysis depending on if the $K^0_S \to \mu^+\mu^-$ candidate was responsible for the trigger or some other particle in the underlying event was. The first type of candidates are called TOS (Trigger On Signal), while the second TIS (Trigger Independent of Signal).

After the trigger and selection, further background discrimination is achieved by the use of a Boosted Decision Tree (BDT), as in the $B^0_{(s)} \to \mu^+\mu^-$ case. However, the BDT is trained here using data, which is splitted in two samples so that the BDT applied in each half has been trained in the complementary half. With all this, four independent samples (two TIS and two TOS) are used for the search for $K^0_S \to \mu^+\mu^-$.

The signal candidates surviving the selection are classified in bins of the BDT, and a limit is computed using the CL_s method once again. The background expectation is obtained from the mass sidebands, and is composed mainly by combinatorial background and $K^0_S \to \pi^+\pi^-$ with both pions misidentified as muons.

Finally, the normalisation is done in bins of BDT (contrary to $B^0_{(s)} \to \mu^+\mu^-$), using $K^0_S \to \pi^+\pi^-$ in an independent way for the TIS and TOS analyses. To this end, for the first $K^0_S \to \pi^+\pi^-$ TIS events are employed, whereas for the second the simultaneous use of $K^0_S \to \pi^+\pi^-$ from physics trigger lines and from minimum bias lines[1] is required. The ratio of reconstruction and selection efficiencies are obtained in a similar way to the $B^0_{(s)} \to \mu^+\mu^-$ case.

[1] heavily prescaled lines in which the trigger takes a random decision.

It is also important to say that no mass binning is used in this search.

As in the $B^0_{(s)} \to \mu^+\mu^-$ case, this search was blinded, and the signal region was not looked at until the analysis was considered mature enough.

8.5.1 Selection, trigger and BDT

The selection of the $K^0_S \to \mu^+\mu^-$ and $K^0_S \to \pi^+\pi^-$ candidates is the same, except for the PID cuts in the first channel, and a prescale in the second (imposed for technical reasons, to avoid a too high rate).

As explained, the understanding of the trigger effects turns out to be crucial in this analysis. Because of its very low efficiency, an important fraction of the selected candidates are TIS, with very different properties to those being TOS. The properties of the TIS signal can be well reproduced in data using $K^0_S \to \pi^+\pi^-$ TIS, but a different strategy has to be used in the TOS case. In this way, a set of tight PID cuts are defined to select $K^0_S \to \pi^+\pi^-$ candidates with both pions misidentified as pions as similar as possible to $K^0_S \to \mu^+\mu^-$. This sample is used as a proxy for the signal in the analysis.

The BDT is also independent in the TIS and TOS cases, using slightly different variables. As a particular feature here, an important fraction of the sample comes from material interactions, so information to veto these is added to the BDT. Studies on the best BDT configuration and best choice of variables have been performed. The BDT is trained using $K^0_S \to \pi^+\pi^-$ TIS and $K^0_S \to \pi^+\pi^-$ with double misidentification (misID) TOS (with the just described PID cuts) as signal and the $K^0_S \to \mu^+\mu^-$ right mass sideband as background. As explained, this forces the split of the TIS and TOS samples in two, having as a final result 4 independent analyses.

8.5.2 Specific backgrounds

Apart from the combinatorial background, $K^0_S \to \pi^+\pi^-$ with double muon misID is the most relevant background for $K^0_S \to \mu^+\mu^-$. The excellent LHCb invariant mass resolution moves this background to the left sideband, but it still enters in the $K^0_S \to \mu^+\mu^-$ signal region with a tail that has to be parameterised. Figure 8.4 shows the $K^0_S \to \mu^+\mu^-$ invariant mass spectra of the selected candidates, with the signal region blinded.

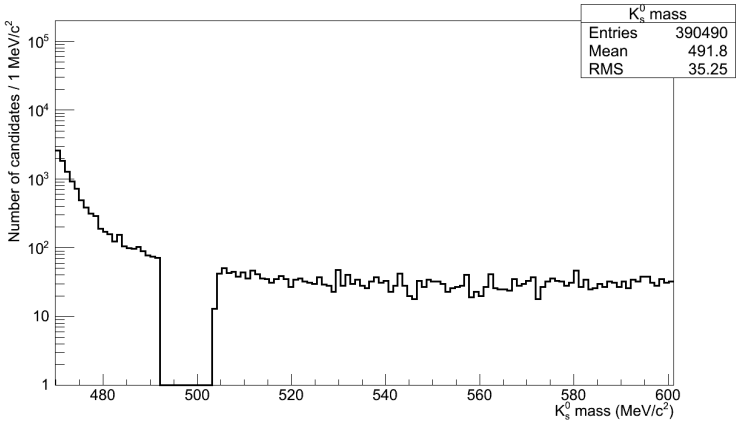

Figure 8.4: Mass distribution of $K^0_S \to \mu^+\mu^-$ after selection. The signal region was blinded in this stage of the analysis.

The best PDF for describing the double misID tail was determined in MC. Then, this PDF was applied in data in each of the BDT bins, added linearly to an exponential PDF to account for the combinatorial background, which is dominant in the right hand side mass sideband.

$K^0_L \to \mu^+\mu^-$, $K^0_L \to \mu^+\mu^-\gamma$, $K^0_L \to \pi^+\mu^-\bar{\nu}_\mu$, $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ and $K^0_S \to \mu^+\mu^-\gamma$ have also been checked to be unharmful for this version of the analysis, but relevant in future searches.

8.5.3 Normalisation

The normalisation to $K^0_S \to \pi^+\pi^-$ is also independent for the TIS and TOS categories, and is performed in bins of BDT. Again, the normalisation is separated in reconstruction, selection and trigger ratios of efficiencies. However, for convenience, the PID is detached here from the rest of selection factors. As mentioned, different strategies are used for the number of candidates in the TIS and TOS cases. The normalisation can be summarised to:

- Reconstruction and selection: the ratio of efficiencies is determined in MC, using the data-corrected p_T and rapidity spectra of K^0_S.
- PID: the $K^0_S \to \mu^+\mu^-$ PID absolute efficiency is obtained reweighting the muonID efficiency curves obtained in data with the $K^0_S \to \pi^+\pi^-$ p, p_T spectra also from data.
- Trigger: for the TIS case, the ratio is simply 1, this verified in MC. For the TOS, the absolute $K^0_S \to \mu^+\mu^-$ TOS efficiency is obtained as in the PID case, reweighting the TOS efficiency curves obtained in data.
- Number of candidates: for the TIS case, the number of $K^0_S \to \pi^+\pi^-$ TIS is obtained. For the TOS, the total number of selected candidates before any trigger is used. This is achieved using the minimum bias sample and determining the effective prescale of this line throughout the year 2011.

8.5.4 Results

No significant signal excess has been found in the search for $K^0_S \to \mu^+\mu^-$. Figure 8.5 shows the background expected and observation in the most sensitive region of the BDT for the 4 samples of the analysis.

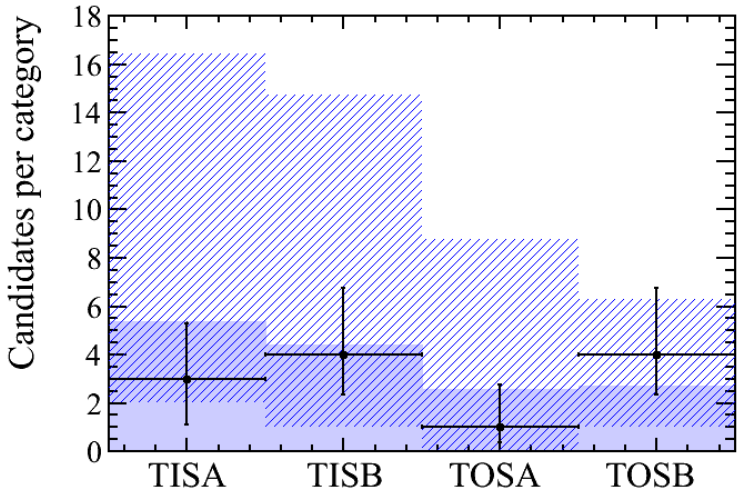

Figure 8.5: Background expected and observation in the most sensitive region of the BDT for the 4 samples of the analysis. The shaded area reflects the error in the expected background.

Using the CL_s method, upper limits on the $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ have been set: $\mathcal{B}(K^0_S \to \mu^+\mu^-) < 11.2 \times 10^{-9}$ ($\mathcal{B}(K^0_S \to \mu^+\mu^-) < 9.0 \times 10^{-9}$) at 95 % CL (90% CL). This upper limit is a factor of 35 better than the previous world best.

9 Resumo

O Modelo Estándar (SM) da física de partículas é a teoría máis acertada para explicar a estrutura e relacións de todas as partículas fundamentais. Estas inclúen os nucleóns e electróns que compoñen os átomos, os fotóns e todo un grupo enorme de novas partículas descubertas ao longo do século XX e comezos do XXI.

A pesares do seu éxito, aínda hai certas medidas experimentais que o SM non é capaz de explicar completamente. Tendo isto en conta, téñense proposto diversos modelos de nova física (NP) para resolver estes problemas

Experimentos como o LHC, no CERN, foron creados para poñer a proba as predicións do SM e dos modelos de NP. O LHC fai colidir protóns entre si a unha enerxía de $\sqrt{s} = 7$ TeV, para que logo varios detectores rexistren o resultado destas colisións. Entre estes detectores está LHCb, deseñado para estudar a violación CP e decaementos raros como $B^0_{(s)} \to \mu^+\mu^-$ e $K^0_{\rm S} \to \mu^+\mu^-$.

9.1 Aspectos teóricos de $B^0_{(s)} \to \mu^+\mu^-$ e $K^0_{\rm S} \to \mu^+\mu^-$

$B^0_{(s)} \to \mu^+\mu^-$ e $K^0_{\rm S} \to \mu^+\mu^-$ son decaementos moi raros, con cocientes de ramificación (tamén chamados fraccións de desintegración ou $\mathcal{B}$), no SM suprimidos ao nivel de $10^{-9} - 10^{-12}$. Estes modelos son, porén, sensíbeis a NP, de xeito que o seu $\mathcal{B}$ podería medrar (ou mesmo estar máis suprimido) con respecto ás predicións do SM. Xa que logo, a medida dun $\mathcal{B}$ significativamente distinto do do SM sería unha proba de física alén o SM.

Os modos de desintegración $B^0_{(s)} \to \mu^+\mu^-$ e $K^0_{\rm S} \to \mu^+\mu^-$ son correntes neutras de cambio de sabor, polo que están prohibidas no SM a nivel de árbore. Os diagramas a través dos cales estes decaementos se producen inclúen *lazos* e *caixas*, nos cales poden entrar partículas que non pertencen ao SM e cambiar, como se comentou, o valor real das fraccións de desintegración con respecto da predición do SM. Estas canles son, polo tanto, axeitadas para unha procura indirecta de NP, coa vantaxe de acadar escalas enerxéticas máis altas que nas procuras directas.

Os decaementos $B^0_{(s)} \to \mu^+\mu^-$ atopánse entre os favoritos na busca de NP no sector de sabor. Os seus $\mathcal{B}$ teñen sido amplamente estudados, e existen predicións exactas dentro do SM $\mathcal{B}(B^0 \to \mu^+\mu^-)|_{SM} = (1{,}1 \pm 0{,}1) \times 10^{-10}$ e $\mathcal{B}(B^0_s \to \mu^+\mu^-)|_{SM} = (3{,}2 \pm 0{,}3) \times 10^{-9}$. Como se comentou, estas fraccións de desintegración son moi sensíbeis á presenza de NP, e polo tanto existe moita literatura na teoría onde se describe como poderían verse estes $\mathcal{B}$ afectados dependendo de espazo de fases da NP. O cadro 9.1 resume algúns destes resultados no caso de $B^0_s \to \mu^+\mu^-$, que ten o $\mathcal{B}$ maior no SM (e polo tanto é máis doado que sexa posto a proba nos experimentos).

Escenario	Apuntaría a
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \gg$SM	Grandes aumentos na NP no sector escalar, SUSY con tan β alto
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \neq$SM	SUSY, EDs, LHT, TC2
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \sim$SM	descartar rexións do parámetro de fases que predín trocos notábeis con respecto do SM
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \ll$SM	NP no sector escalar, pero MSSM case descartado ou moi improbábel. O NMSSM sería un bo candidato

Cadro 9.1: Implicacións para a NP de diferentes valores do $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$

O decaemento $K_S^0 \to \mu^+\mu^-$ propúxose como unha maneira novidosa de poñer a proba o SM empregando o potencial do experimento LHCb. No SM, $K_S^0 \to \mu^+\mu^-$ está basicamente dominado por física de longa distancia, cunha contribución de física de distancias curtas moito menor. Porén, esta parte de distancias curtas é sensíbel a NP, de xeito que podería converterse mesmo en dominante dependendo do espazo de fases da NP, facendo medrar o $\mathcal{B}$ da canle case unha orde de magnitude por riba da predición do SM, que é $\mathcal{B}(K_S^0 \to \mu^+\mu^-)|_{SM} = (5{,}1 \pm 1{,}5) \times 10^{-12}$. Tense suxerido $K_S^0 \to \mu^+\mu^-$ como un xeito de buscar novos escalares lixeiros e tamén de establecer límites independentes do modelo na fase de violación CP da amplitude $s \to dl^+l^-$.

9.2 O experimento LHCb no LHC

LHCb é un dos catro grandes detectores do experimento LHC no CERN. LHCb é un detector de propósito específico: estudar a violación CP e os decaementos raros, sobre todo a partir de decaementos de partículas co quark b.

LHCb deseñouse para funcionar a unha luminosidade instantánea menor que a dos outros experimentos do LHC. A razón disto é facilitar a correcta identificación dos vértices primario (PV) e secundario (SV), dado que isto é esencial para os obxectivos físicos de LHCb.

9.2.1 O detector LHCb

LHCb (figura 9.1) é un espectrómetro de brazo único que cubre a zona dianteira cunha cobertura angular desde aproximadamente 10 mrad a 300 (250) mrad no plano perpendicular ao campo magnético (paralelo ao campo magnético). Os principais elementos de LHCb son:

- un imán composto por un dipolo a temperatura ambiente que proporciona un campo integrado de 4 T·m,
- un sistema localizador de vértices (VELO),
- un sistema de trazado composto polo Tracker Turiciensis (un detector de microtiras de silicio, TT) en fronte do imán, e tres estacións de trazado despois do imán, formadas por microtiras de silicio na parte interior (IT) e por tubos de gas na parte exterior (OT),
- dous contadores de Aneis Cherenkov (RICH1 e RICH2), para acadar unha excelente separación p–K no rango de momento de 2 a 100 GeV/c,
- un sistema de calorímetros formado por unha capa cintiladora previa (SPD/PS), un calorímetro electromagnético (ECAL) e un calorímetro hadrónico (HCAL),
- un sistema de detección de muóns composto por MWPC.

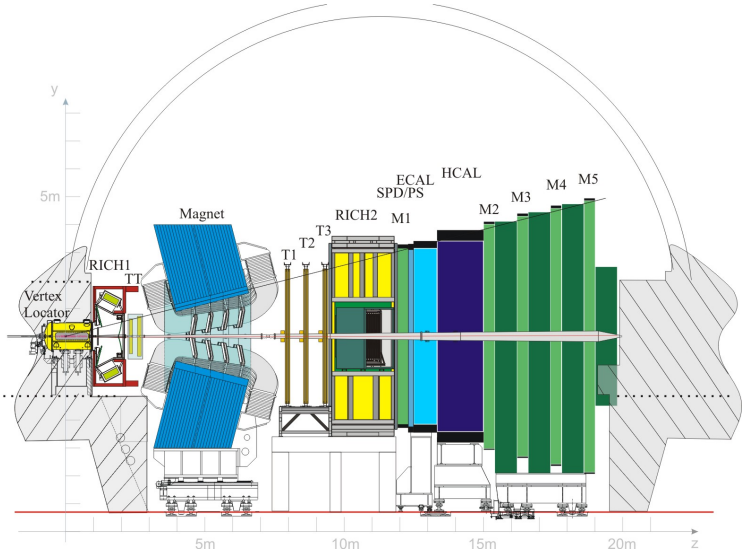

Figura 9.1: Vista lateral de LHCb

9.2.2 O fluxo de traballo en LHCb

Os datos tomados por LHCb a partir das colisións protón–protón do LHC, a un ritmo de varios millóns de sucesos por segundo, precisan ser seleccionados, preparados e distribuídos da maneira o máis eficiente posíbel para a súa posterior análise. Este proceso resúmese a continuación:

1. A cantidade de datos xerados nas colisións do LHC é demasiada alta para ser directamente almaceada. Tendo isto en conta, faise preciso unha fase inicial de selección rápida que permita descartar a maioría de sucesos que non son interesantes para as análises físicas. Esta selección inicial denomínase *trigger*, e é común a todos os experimentos de física de altas enerxías. En particular, o trigger de LHCb no ano 2011 encargouse de reducir a frecuencia dos datos de ∼15 MHz a ∼3 kHz.

2. Os datos seleccionados polo trigger son transformados por unha serie de algoritmos matemáticos nun conxunto de trazas e vértices. A reconstrución de trazas e vértices faise cos detectores de trazado (incluído o VELO). A información dos detectores de identificación de partículas (PID) engádese despois para determinar a natureza das trazas, distinguindo, por exemplo, muóns doutras partículas.

3. Unha vez todos os sucesos que pasaron o trigger teñen sido reconstruídos, faise preciso separalos de acordo ao seu contido físico. Isto acádase seleccionando os diferentes decaementos segundo as súas características específicas. Por exemplo, $B^0_{(s)} \to \mu^+\mu^-$ ou $K^0_S \to \mu^+\mu^-$, co mesmo estado final, teñen diferente masa e tempo de vida da nai, separación agardada das fillas con respecto do PV,...

4. Os datos despois do trigger, reconstrución e selección son distribuídos a unha serie de centros de computación espallados arredor do mundo. Con isto, o acceso aos datos convírtese en máis doado, e a súa análise pode ser distribuída. Todo isto conséguese grazas ao sistema do *Grid*.

5. A última ferramenta precisa para producir resultados físicos é o uso de datos simulados mediante Monte Carlo (MC). Isto é crucial para entender diversos efectos e nesgos creados polo grupo de procesos que se vén de describir. A produción de datos MC o máis similares posíbel aos reais é un dos maiores esforzos de todos os experimentos de física de altas enerxías.

As análises desta tese están baseadas nos datos tomados por LHCb durante 2011 a partir das colisións protón–protón do LHC a unha enerxía no centro de masas de $\sqrt{s} = 7$ TeV. A luminosidade neste período mantívose no rango $3-3{,}5\times10^{32}$ cm^{-2} s^{-1} . O total de luminosidade integrada corresponde a ~ 1.02 fb^{-1}.

9.3 Procura de decaementos raros en LHCb

Algúns dos resultados máis senlleiros de LHCb ata o de agora veñen na área dos decaementos raros. Os decaementos raros inclúen varias análises con características experimentais comúns, o que fai sinxelo compartir ferramentas e experiencia en como resolver problemas entre todas elas. As buscas de decaementos raros en LHCb divídense tipicamente en transicións con pengüíns electrofebles $b \to d\mu^+\mu^-$ e $b \to s\mu^+\mu^-$, decaementos radiativos e decaementos moi raros.

As transicións con pengüíns electrofebles como $b \to d\mu^+\mu^-$ e $b \to s\mu^+\mu^-$ producen unha serie de observábeis sensíbeis á NP. En particular, no caso de $B^0 \to K^{*0}\mu^+\mu^-$ estes observábeis son accesíbeis mediante o uso das distribucións angulares dos produtos do decaemento, e son dependentes da masa invariante do par de muóns. Esta dependencia esta ben predicida no SM e é distinta noutros modelos. LHCb ten medido estes observábeis cos datos do 2011, e os resultados obtidos son os mellores do mundo, compatíbeis ademais co SM. Outro observábel do mesmo tipo é a medida da asimetría de isoespín nos modos $B \to K\mu^+\mu^-$ e $B \to K^{*0}\mu^+\mu^-$. Empregando os datos de 2011, LHCb mediu unha asimetría compatíbel con 0 no caso de $B \to K^{*0}\mu^+\mu^-$ (de acordo coa predición do SM), pero significativemente por baixo para o $B \to K\mu^+\mu^-$. Mesmo se aínda non hai unha predición no SM para o segundo modo, agárdase que esta sexa tamén 0. Como exemplo final do traballo coas transicións con pengüíns electrofebles, LHCb ten descuberto o decaemento dun mesón B máis raro ata o de agora, $B^+ \to \pi^+\mu^+\mu^-$, cun $\mathcal{B}$ no nivel de $\sim 2{,}4 \times 10^{-8}$.

Os decaementos radiativos con b tamén poderían amosar signos de física alén do SM, dado que a NP pode introducir efectos medíbeis na dinámica das transicións radiativas. Un exemplo de observábel sensíbel a estes efectos de NP é a asimetría CP directa en $b \to s\gamma$ usando a canle $B^0 \to K^{*0}\gamma$. Empregando o 1 fb^{-1} recollido por LHCb durante 2011, tense medido unha asimetría CP consistente coa predición do SM, moi próxima a 0.

A etiqueta de "decaementos moi raros" refírese a modelos extremadamente suprimidos no SM, con posíbeis aumentos (ou mesmo maiores supresións) en modelos de NP, de xeito que unha observación directa ou límites superiores no $\mathcal{B}$ poidan poñer límites no espazo fásico da NP. $B^0_{(s)} \to \mu^+\mu^-$ e $K^0_S \to \mu^+\mu^-$ son exemplos de decaementos moi raros. Ademais destes, LHCb ten buscado outros decaementos moi raros empregando os datos de 2011. Exemplos son os decaementos con violación do sabor leptónico $\tau^- \to \mu^+\mu^-\mu^-$ e $B^- \to X^+\mu^-\mu^-$, e tamén outras canles como $B^0_{(s)} \to \mu^+\mu^-\mu^+\mu^-$ e $D^0 \to \mu^+\mu^-$. LHCb non ten atopado ningún exceso significativo de sinal nestes modos, pero ten establecido mellores límites superiores mundiais (ou preto destes) nas fraccións de desintegración de todos eles.

9.4 $B^0_{(s)} \to \mu^+\mu^-$

Debido ao seu interese, os decaementos $B^0_{(s)} \to \mu^+\mu^-$ teñen sido procurado por moitos experimentos anterioremente. Os mellores límites superiores ao comezo de 2012 eran os atopados por LHCb, empregando arredor do 30 % dos datos de 2011: $\mathcal{B}(B^0 \to \mu^+\mu^-) < 3{,}6 \times 10^{-9}$ e $\mathcal{B}(B^0_s \to \mu^+\mu^-) < 1{,}6 \times 10^{-8}$, as dúas ao 95 % de nivel de confianza (CL).

Para a busca destas canles en LHCb tense deseñado unha selección moi eficiente para desfacerse da maioría do fondo. O número de sucesos observado compárase despois co agardado de fondo e sinal en rangos de dúas variábeis independentes, a masa invariante e o resultado dun discriminante multi-variábel, o BDT. A probabilidade de que os sucesos do sinal e do fondo teñan un determinado valor do BDT extráese dos datos empregando candidatos de $B^0_{(s)} \to h^+h'^-$ para o sinal e as bandas laterais de masa de $B^0_{(s)} \to \mu^+\mu^-$ para o fondo

combinatorio. A distribución en masa dos sinais descríbese mediante unha función "Crystal Ball", cuns parámetros que tamén se extraen dos datos. Para o fondo, esta distribución de masa extráese novamente das bandas laterais de $B^0_{(s)} \to \mu^+\mu^-$. O número agardado de sucesos de sinal para unha determinada hipótese da fracción de desintegración obtense normalizando a canles con fraccións de desintegración coñecidas: $B^+ \to J/\psi\, K^+$, $B^0_s \to J/\psi\, \phi$ e $B^0 \to K^+\pi^-$.

Tamén se ten comprobado que diversos fondos específicos son desprezábeis para a análise. A única excepción é $B^0_{(s)} \to h^+h'^-$ no cal os dous hadróns son tomados por muóns. Aínda que esta contribución é pequena grazas aos cortes en PID, tamén se ten parametrizado nos rangos de masa e BDT.

Unha vez que os candidatos observados e agardados de sinal e fondo están dispoñíbeis en cada rango do espazo BDT–masa invariante, a súa compatibilidade é calculada empregando o método $\mathrm{CL_s}$. Con isto, e dado que o número agardado de sinal é propocional á hipótese de fracción de desintegración, pódense establecer límites superiores nesta a un nivel de confianza determinado.

É importante dicir tamén que esta análise é "cega", isto é, a rexión de masa non se olla ata que a análise se considere completa.

9.4.1 Selección e BDT

A seleccion do $B^0_{(s)} \to \mu^+\mu^-$ foi deseñada para ser similar nas canles de sinal e control, para así reducir os erros sistemáticos cando se realiza a normalización e calibración. En particular, a única diferenza entre as seleccións de $B^0_{(s)} \to \mu^+\mu^-$ e $B^0_{(s)} \to h^+h'^-$ son os cortes de PID aplicados aos muóns no primeiro caso. Tamén é importante dicir que un BDT inicial que emprega variábeis sinxelas aplícase xa a nivel de selección. Este BDT denomínase BDTS. O BDTS foi deseñado para proporcionar as mesmas eficiencias nas canles de normalización e sinal.

O BDT empregado para clasificar os sucesos que sobreviven á selección estudouse en detalle, para así poder buscar a mellor parametrización e elección de variábeis que produzan o mellor resultado. Deste xeito, a elección final incluiu 9 variábeis diferentes que recollen propiedades dos muóns e do mesón B. O adestramento do BDT faise empregando $B^0 \to \mu^+\mu^-$ do MC para o sinal e $b\bar{b} \to \mu\mu X$ tamén do MC para o fondo.

Tamén se levou a cabo un estudo para definir os rangos en BDT (e masa) que lle proporcionan á análise unha maior sensibilidade.

9.4.2 Fondos específicos

O fondo máis relevante para esta análise é o combinatorio, composto principalmente por $b\bar{b} \to \mu\mu X$, isto é, pares de quark b que decaen independentemente a un muón.

O fondo específico máis importante é o $B^0_{(s)} \to h^+h'^-$ no cal os dous hadróns son tomados por muóns (misID). A contribución deste fondo foi obtida medindo a frecuencia de misID a partir dos kaóns e pións en $D^0 \to K^+\pi^-$ en función do seu p, p_T e aplicando o mapeado obtido ao espazo fásico do $B^0_{(s)} \to h^+h'^-$, obtido no MC. O total agardado deste fondo obtense a partir do $B^0_{(s)} \to h^+h'^-$ total nos datos de 2011. A súa distribución de BDT e masa tamén a partir do MC.

Tense visto ademais que outros fondos específicos como $B^+_c \to J/\psi(\to \mu^+\mu^-)\mu^+\nu$, $B^+ \to \pi^+\mu^+\mu^-$ e $B_s \to \mu^+\mu^-\gamma$ son desprezábeis para esta busca.

9.4.3 Calibración do BDT e da masa invariante

A función de densidade de probabilidadae (PDF) do sinal obtense a partir de sucesos $B^0_{(s)} \to h^+h'^-$ nos cales o mesón B non foi responsábel do trigger. Así, a PDF do BDT obtense empregando diferentes funcións de axuste, coa diferenza medida tomada como erro sistemático. Un exemplo destes axustes para os dous últimos rangos de BDT pódese atopar na figura 9.2. Despois disto, a PDF producida corríxese para ter en conta o efecto do trigger e dos cortes de identificación de muóns (muonID).

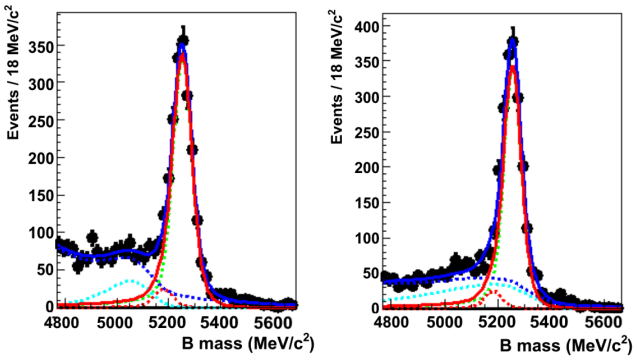

Figura 9.2: Axuste ao espectro de masas de $B^0_{(s)} \to h^+h'^-$ para obter a calibración do BDT. Este exemplo amosa os dous últimos rangos de BDT.

As PDF para a masa do sinal parametrízanse como dúas funcións Crystal Ball. Os valores centrais extráense das mostras exclusivas $B^0 \to K^+\pi^-$ e $B^0_s \to K^+K^-$, mentres que as resolucións mídense interpolando á masa dos mesóns B^0 e B^0_s as resolucións medidas coas resonancias J/ψ, $\Psi(2S)$, $\Upsilon(1S, 2S, 3S)$.

As PDF para a masa invariante e BDT do fondo combinatorio extráense dun axuste das bandas laterais de masa dos dimuóns en cada rango de BDT, empregando unha PDF exponencial e outros modelos como comprobación.

9.4.4 Normalización

A normalización a unha canle cunha fracción de desintegración coñecida permite converter un número de candidatos nun $\mathcal{B}$ sen precisar coñecer a luminosidade absoluta e a sección eficaz $b\bar{b}$. As canles da normalización en LHCb son: $B^+ \to J/\psi\, K^+$, $B^0_s \to J/\psi\, \phi$ e $B^0 \to K^+\pi^-$. Deste xeito, a normalización require medir os cocientes das eficiencias de reconstrución, selección e trigger entre o sinal as canles de calibración mencionadas, así como o cociente de fraccións de produción (dependendo da canle de calibración) e o número de candidatos da calibración:

- Fraccións de produción: úsase o cociente $f_{B^0}/f_{B^0_s}$ medido por LHCb, sempre e cando a normalización implique dous modos con diferentes nais.
- Reconstrución: para a reconstrución emprégase simulación MC, validando logo o resultado nos datos.
- Selección: para os cortes xeométricos, úsase novamente simulación MC, pero cos parámetros "espallados" para que a súa resolución se pareza á dos datos. Para o PID, os cocientes extraénse a partir dos datos combinados co MC.
- Trigger: a eficiencia do trigger para as canles de normalización con muóns obtense directamente nos datos. Para o sinal, pesando a eficiencia medida nos datos co espectro de p e IP de $B^0 \to \mu^+\mu^-$ do MC. No caso de $B^0 \to K^+\pi^-$, a normalización faise cos casos nos cales o mesón B^0 non é responsábel da decisión do trigger, para os cales a eficiencia se mide empregando o decaemento $B^+ \to J/\psi\, K^+$.
- Número de candidatos: o número de candidatos obtense realizando cadanseu axuste ao espectro de masa invariante.

9.4.5 Resultados

A busca de $B^0_{(s)} \to \mu^+\mu^-$ non resultou en ningún exceso significativo de sinal. A figura 9.3 amosa a distribución de masa invariante dos candidatos observados e as contribucións agardadas de fondo e sinal do SM no rango de BDT por riba de 0.5

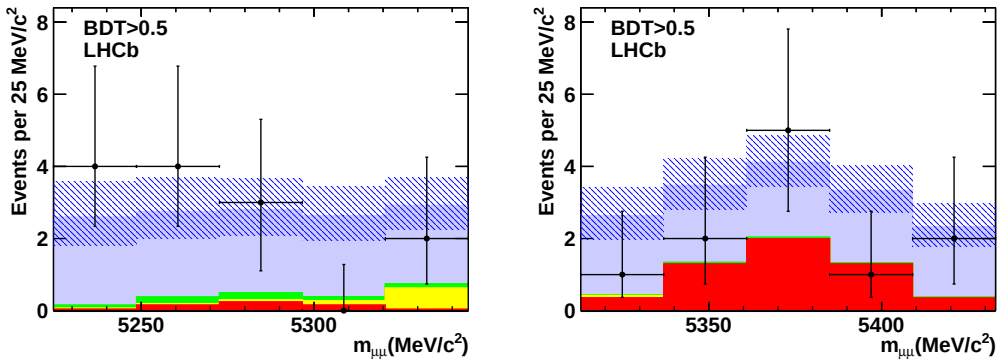

Figura 9.3: Distribución de masa invariante dos candidatos seleccionados de $B^0 \to \mu^+\mu^-$ (esquerda) e $B^0_s \to \mu^+\mu^-$ (dereita) con BDT$>$ 0,5. Amósanse os candidatos observados e as diferentes contribucións de fondo. O efecto dun sinal potencial do SM tamén se amosa. A área sombreada refrexa o erro na suma de todas as contribucións.

Dado que non se atopou ningún exceso significativo de sinal, puidéronse establecer límites superiores nas fraccións de desintegración das dúas canles. Como se explicou, o método de CL_s é o empregado para obter estes límites. Os dous resultados, amosados no cadro 9.2, son os mellores mundiais, e producen importantes límites no espazo fásico da NP.

Canle	Límite ao 95 % de CL	Límite ao 90 % de CL
$B^0 \to \mu^+\mu^-$	$1{,}0 \times 10^{-9}$	$0{,}8 \times 10^{-9}$
$B^0_s \to \mu^+\mu^-$	$4{,}5 \times 10^{-9}$	$3{,}8 \times 10^{-9}$

Cadro 9.2: Límites superiores nos $\mathcal{B}$ dos decaementos $B^0 \to \mu^+\mu^-$ e $B^0_s \to \mu^+\mu^-$ empregando os datos de LHCb do 2011.

9.5 $K^0_S \to \mu^+\mu^-$

O límite superior na $\mathcal{B}(K^0_S \to \mu^+\mu^-)$ é $3{,}2 \times 10^{-7}$ ao 90 % de CL, obtido hai case 40 anos.

A busca de $K^0_S \to \mu^+\mu^-$ é complexa debido ao feito de que LHCb está optimizado para mesóns B, que teñen características distintas ao K^0_S. Como exemplo, a masa dos K^0_S é arredor de 10 veces máis pequena que a do B^0, mentres que voan un factor 50 máis. Isto produce unha reconstrución e un trigger moi distintos.

A análise de $K^0_S \to \mu^+\mu^-$ en LHCb tamén comeza cunha selección eficiente para desfacerse da maioría do fondo. Porén, os grandes nesgos producidos pola baixa eficiencia do trigger obrigan a separar a análise dependendo de se o candidao a $K^0_S \to \mu^+\mu^-$ foi o responsábel do trigger ou se outra partícula no mesmo suceso o foi. Ao primeiro tipo de candidatos chámaselles TOS (trigger no sinal) e ao segundo TIS (trigger independente do sinal).

Despois da selección e o trigger, conséguese máis discriminación empregando un BDT, como no caso do $B^0_{(s)} \to \mu^+\mu^-$. Non obstante, neste caso o BDT adéstrase nos datos, que se dividen en dúas metades de xeito que o BDT aplicado en cada parte teña sido adestrado coa outra metade complementaria. Con todo isto, catro mostras independentes (dúas TIS e dúas TOS) úsanse na busca de $K^0_S \to \mu^+\mu^-$.

Os candidatos de sinal que sobreviven á selección clasifícanse en rangos do BDT, e obtense un límite empregando novamente o método CL_s. O fondo agardado mídese a partir das bandas laterais de masa, e está composto basicamente dun fondo combinatorio e $K^0_S \to \pi^+\pi^-$ no cal os dous pións son tomados por muóns.

Finalmente, a normalización faise en rangos de BDT (contrariamente ao caso de $B^0_{(s)} \to$

$\mu^+\mu^-$), empregando $K^0_S \to \pi^+\pi^-$ dun xeito independente para as análises TIS e TOS. Deste xeito, para as primeiras úsanse sucesos $K^0_S \to \pi^+\pi^-$ TIS, mentres que a para as segundas faise preciso empregar de xeito simultáneo $K^0_S \to \pi^+\pi^-$ de liñas do trigger físicas e de liñas de nesgo mínimo (MB)[1]. Os cocientes de eficiencias de reconstrución e selección obtéñense de xeito similar ás do $B^0_{(s)} \to \mu^+\mu^-$. Tamén é importante dicir que nesta busca non se utilizan rangos de masa.

Como no caso de $B^0_{(s)} \to \mu^+\mu^-$, esta busca é cega, e non se ollou a rexión de sinal ata que se considerou a análise o suficientemente madura.

9.5.1 Selección, trigger e BDT

A selección dos candidatos de $K^0_S \to \mu^+\mu^-$ e $K^0_S \to \pi^+\pi^-$ é a mesma, salvo polos cortes en PID na primeira canle, e pola redución aplicada na segunda (imposta por motivos técnicos, para evitar unha frecuencia demasiado alta).

Como se explicou, entender os efectos do trigger resulta ser crucial para esta análise. Pola súa baixa eficiencia, unha fracción importante dos candidatos seleccionados son TIS, cunhas propiedades diferentes aos que son TOS. As propiedades do sinal TIS poden ser ben reproducidas nos datos empregando $K^0_S \to \pi^+\pi^-$ TIS, pero para os TOS tense que usar unha estratexia distinta. Neste senso, un grupo de cortes de PID máis duros defínense para seleccionar aqueles candidatos $K^0_S \to \pi^+\pi^-$ cos dous pións tomados por muóns o máis parecidos posíbel ao $K^0_S \to \mu^+\mu^-$. Esta mostra emprégase como "representante" do sinal na análise.

O BDT tamén é distinto nos TIS e TOS, empregando de feito variábeis lixeiramente distintas. Como unha característica peculiar cómpre destacar que unha parte importante da mostra provén de interaccións co material, o que fai preciso engadir información para vetalas no BDT. Realizáronse tamén estudos para achar a mellor configuración do BDT e a mellor elección de variábeis. O BDT adéstrase con $K^0_S \to \pi^+\pi^-$ TIS e $K^0_S \to \pi^+\pi^-$ con dobre misID TOS (cos cortes en PID dos que se falou) como sinal e coa banda lateral de masa dereita do $K^0_S \to \mu^+\mu^-$ nos datos como fondo. Como se comentou, isto forza a división das mostras TIS e TOS en dous, tendo como resultado final 4 análises independentes.

9.5.2 Fondos específicos

Ademais do fondo combinatorio, o $K^0_S \to \pi^+\pi^-$ onde os dous pións se toman por muóns é o fondo máis relevante para $K^0_S \to \mu^+\mu^-$. A excelente resolución en masa invariante de LHCb permite mover este fondo á banda lateral de masa esquerda, pero aínda entra na rexión de sinal do $K^0_S \to \mu^+\mu^-$ cunha cola que ten que ser parametrizada. A figura 9.4 amosa o espectro de masa invariante de $K^0_S \to \mu^+\mu^-$ dos candidatos seleccionados, coa rexión de sinal "cegada".

A mellor PDF para describir a cola do dobre misID determinouse no MC. Logo diso, esta PDF aplicouse nos datos en cada rango de BDT, engandindo linearmente unha PDF exponencial para ter en conta tamén o fondo combinatorio, que é o dominante na banda lateral de masa dereita.

Tamén se comprobou que outros fondos como $K^0_L \to \mu^+\mu^-$, $K^0_L \to \mu^+\mu^-\gamma$, $K^0_L \to \pi^+\mu^-\bar{\nu}_\mu$, $K^0_S \to \pi^+\mu^-\bar{\nu}_\mu$ e $K^0_S \to \mu^+\mu^-\gamma$ non son perigosos para esta versión da análise, pero poderían selo en futuras buscas.

9.5.3 Normalización

A normalización a $K^0_S \to \pi^+\pi^-$ tamén é independente para as categorías TIS e TOS, e lévase a cabo en rangos de BDT. Novamente, esta normalización sepárase en cocientes de eficiencias de reconstrución, selección e trigger. Porén, por conveniencia, o PID sepárase neste caso do resto dos factores da selección. Como se mencionou, empréganse diferentes estratexias para o número de candidatos nos casos TIS e TOS. A normalización resúmese en:

[1] liñas moi reducidas nas cales o trigger toma unha decisión aleatoria.

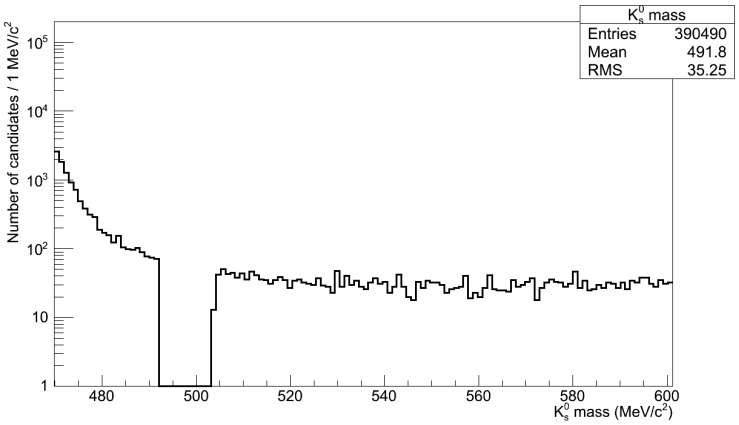

Figura 9.4: Distribución de masa de $K^0_S \to \mu^+\mu^-$ despois da selección. A rexión de sinal estaba "cegada" neste período da análise.

- Reconstrución e selección: o cociente de eficiencias determínase no MC, empregando o p_T e rapidez dos K^0_S corrixidas cos datos.
- PID: a eficiencia absoluta de PID do $K^0_S \to \mu^+\mu^-$ obtense repesando as curvas de eficiencia do muonID obtidas nos datos co espectro de p, p_T obtido tamén nos datos.
- Trigger: para o caso dos TIS, o cociente é simplemente 1, o cal se verificou nos datos. Para o TOS, obtense a eficiencia absoluta de $K^0_S \to \mu^+\mu^-$ TOS dun xeito similar á do PID, repesando as curvas de eficiencia TOS obtidas cos datos.
- Número de candidatos: para o caso TIS, obtense directamente o número de $K^0_S \to \pi^+\pi^-$ TIS. Para os TOS, úsase o número total de candidatos seleccionados antes do trigger. Isto acádase a través da mostra MB e determinando a redución efectiva desta liña ao longo do ano 2011.

9.5.4 Resultados

Non se ten atopado ningún exceso significativo de sinal na busca de $K^0_S \to \mu^+\mu^-$. A figura 9.5 amosa o fondo agardado e observado na rexión de BDT máis sensíbel do BDT das catro mostras da análise.

Empregando o método CL_s, establecéronse límites superiores no $\mathcal{B}(K^0_S \to \mu^+\mu^-)$: $\mathcal{B}(K^0_S \to \mu^+\mu^-) < 11{,}2 \times 10^{-9}$ ($\mathcal{B}(K^0_S \to \mu^+\mu^-) < 9{,}0 \times 10^{-9}$) ao 95 % de CL (90 % de CL). Este límite superior é un factor 35 mellor que o previo mellor do mundo.

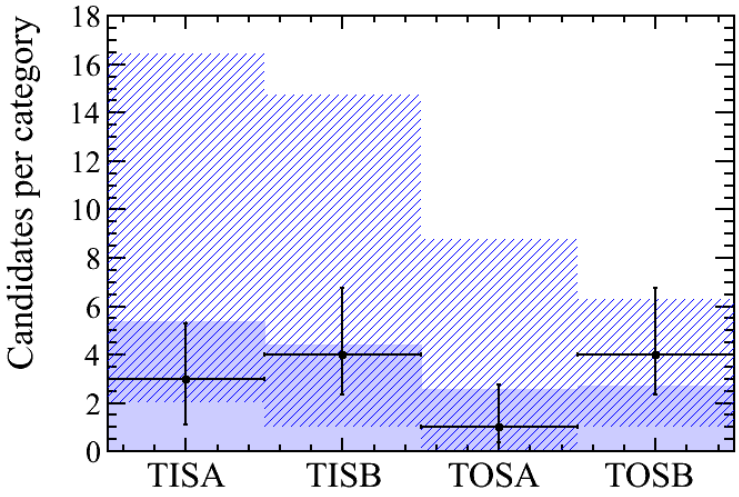

Figura 9.5: Fondo agardado e observación na rexión máis sensíbel do BDT nas 4 mostras da análise. A área sombreada refrexa o erro no fondo agardado.

Bibliography

[1] D. Martinez, *Study of the very rare decay $B_s^0 \to \mu^+\mu^-$ in LHCb*, *CERN-THESIS-2010-068* (2010).

[2] `http://dorigo.wordpress.com/2007/11/10/the-goldstone-theorem-for-real-dummies/`.

[3] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, *Phys. Lett.* **10** (1963).

[4] M. Kobayashi and T. . Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973).

[5] **Particle Data Group** Collaboration, K. Nakamura *et. al.*, *Review of particle physics*, *J. Phys. G* **G37** (2010) 075021.

[6] T. Inami and C. S. Lim, *Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes $K_{\rm L}^0 \to \mu\mu$, $K^+ \to \pi^+\nu\bar{\nu}$ and $K^0 \leftrightarrow \bar{K}^0$*, *Prog. Theor. Phys.* **65** (1981).

[7] K. G. Wilson, *Non-Lagrangian Models of Current Algebra*, *Phys. Rev.* **179** (1969).

[8] E. Fermi, *Tentativo di una teoria dell'emissione dei Raggi Beta*, *Ricerca Scientifica* **4(2)** (1933).

[9] C. Bobeth *et. al.*, *QCD corrections to $\bar{B} \to X_{d,s}\nu\bar{\nu}, B_{d,s} \to l^+l^-, K \to \pi\nu\bar{\nu}$ and $K_{\rm L}^0 \to \mu^+\mu^-$ in the MSSM*, *Nucl. Phys. B* **630** (2002).

[10] J. Virto, *Topics in Hadronic B Decays*, `arXiv:0712.3367`.

[11] A. K. Alok and S. U. Sankar, *New physics upper bound on the branching ratio of $B_s \to \ell^+\ell^-$*, *Phys.Lett.* **B620** (2005) 61–68, [`hep-ph/0502120`].

[12] **LHCb** Collaboration, R. Aaij *et. al.*, *Determination of f_s/f_d for 7 TeV pp collisions and a measurement of the branching fraction of the decay $B_d \to D^-K^+$*, `arXiv:1106.4435`.

[13] A. J. Buras, *Relations between $\Delta M_{s,d}$ and $B_{d,s} \to \mu^+\mu^-$ in models with minimal flavour violation*, *Phys.Lett. B* **566** (2003).

[14] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, *On the Standard Model prediction for $\mathcal{B}(B_{s,d} \to \mu^+\mu^-)$*, `arXiv:1208.0934`.

[15] D. Melikhov and N. Nikitin, *Rare radiative leptonic decays $B_{d;s} \to l^+l^-\gamma$*, *Phys. Rev. D* **70** (2004).

[16] A. Petrov, *Rare leptonic radiative Bs decays*, *Workshop: Implications of LHCb measurements and future prospects* (2012). `https://indico.cern.ch/getFile.py/access?contribId=11&sessionId=1&resId=0&materialId=slides&confId=176118`.

[17] K. J. Healey *et. al.*, *Soft Photon Contributions to $B_s \to \mu\bar{\mu}$*, *Pheno 2012* (2012). `https://indico.cern.ch/getFile.py/access?contribId=134&sessionId=3&resId=0&materialId=slides&confId=156307`.

[18] K. D. Bruyn *et. al.*, *On Branching Ratio Measurements of B_s^0 Decays*, `arXiv:1204.1735`.

[19] K. D. Bruyn *et. al.*, *A New Window for New Physics in $B_s^0 \to \mu^+\mu^-$*, `arXiv:1204.1737`.

[20] **LHCb** Collaboration, R. Aaij *et. al.*, *Tagged time-dependent angular analysis of $B_s^0 \to J/\psi\,\phi$ decays at LHCb*, *LHCb-CONF-2012-002* (2012).

[21] G. Ecker and A. Pich, *The Longitudinal muon polarization in $K_L \to \mu^+\mu^-$*, *Nucl.Phys.* **B366** (1991) 189–208.

[22] G. Isidori and R. Unterdorfer, *On the short distance constraints from $K_{L,S} \to \mu^+\mu^-$*, *JHEP* **0401** (2004) 009, [`hep-ph/0311084`].

[23] G. D'Ambrosio, G. Ecker, G. Isidori, and H. Neufeld, *Radiative nonleptonic kaon decays*, `hep-ph/9411439`.

[24] G. A. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, *Rev. Mod. Phys.* **68** (1996).

[25] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, *Phys. Rev. Lett.* **51** (1983).

[26] F. Zwicky, *Spectral displacement of extra galactic nebulae*, *Helv. Phys. Acta* **6** (1933).

[27] V. C. Rubin and J. W. K. Ford, *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, *Astrophys. J.* **159** (1970).

[28] V. C. Rubin, N. Thonnard, and W. K. F. Jr., *Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R = 4 kpc) to UGC 2885 (R = 122 kpc)*, *Astrophys. J.* **238** (1980).

[29] D. Clowe *et. al.*, *A direct empirical proof of the existence of dark matter*, *Astrophys. J.* **648** (2006).

[30] Y. B. Zel'dovich and I. D. Novikov, *The Hypothesis of Cores Retarded during Expansion and the Hot Cosmological Model*, *Sov. Astron.* **10** (1966).

[31] P. Ivanov, P. Naselsky, and I. Novikov, *Inflation and primordial black holes as dark matter*, *Phys. Rev. D* **52** (1994).

[32] R. Peccei and H. Quinn, *CP Conservation in the Presence of Pseudoparticles*, *Phys. Rev. Lett.* **38** (1977).

[33] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978).

[34] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978).

[35] J. L. Feng, *Dark Matter Candidates from Particle Physics and Methods of Detection*, `arXiv:1003.0904`.

[36] M. Maltoni *et. al.*, *Status of global fits to neutrino oscillations*, *New J.Phys.* **6** (2004).

[37] J.Ellis *et. al.*, *New parametrization of the seesaw mechanism and applications in supersymmetric models*, *Phys. Rev. D* **66** (2002).

[38] A. Sakharov, *Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe*, *Pisma Zh.Eksp.Teor.Fiz.* **5** (1967) 32–35.

[39] **Muon g-2** Collaboration, G. W. Bennett *et. al.*, *Measurement of the Negative Muon Anomalous Magnetic Moment to 0.7 ppm*, *Phys. Rev. Lett.* **92** (2004).

[40] F. Jegerlehner and A. Nyffeler, *The muon g-2*, *Phys. Rept.* **477** (2009).

[41] **LHCb** Collaboration, R. Aaij *et. al.*, *Evidence for CP violation in time-integrated $D^0 \to h^- h^+$ decay rates*, *Phys. Rev. Lett.* **108** (2012) [arXiv:1112.0938].

[42] **BaBar** Collaboration, J. P. Lees *et. al.*, *Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays*, arXiv:1205.5442.

[43] A. Pomarol, *Beyond the Standard Model*, arXiv:1202.1391.

[44] C. A. Baker *et. al.*, *An improved experimental limit on the electric dipole moment of the neutron*, *Phys. Rev. Lett.* **97** (2006) [0602020].

[45] A. J. Buras and J. Girrbach, *BSM MODELS FACING THE RECENT LHCb DATA: A FIRST LOOK*, *Phys. Rev. Lett.* (2012) [arXiv:1204.5064].

[46] G. Colangelo, E. Nikolidakis, and C. Smith, *Supersymmetric models with minimal flavour violation and their running*, *Eur. Phys. J. C* **59** (2009).

[47] G. D'Ambrosio *et. al.*, *Minimal Flavour Violation: an effective field theory approach*, *Nucl.Phys. B* **645** (2002).

[48] A. J. Buras, P. Gambino, M. Gorbahn, S. Jager, and L. Silvestrini, *Universal Unitarity Triangle and Physics Beyond the Standard Model*, *Phys.Lett. B* **500** (2001).

[49] A. J. Buras, *Minimal Flavour Violation*, *Acta Phys. Polon. B* **34** (2003).

[50] M. Blanke, A. J. Buras, D. Guadagnoli, and C. Tarantino, *Minimal Flavour Violation Waiting for Precise Measurements of ΔM_s, $S_{\psi\phi}$, A^s_{SL}, $|V_{ub}|$, $\gamma and B^0_{s,d} \to mu^+mu^-$*, *JHEP* **0610** (2006).

[51] T. Lee, *A Theory of Spontaneous T Violation*, *Phys. Rev. D* **8** (1973).

[52] L. J. Hall and M. Wise, *Flavor changing Higgs boson couplings*, *Nucl. Phys. B* **187** (1981).

[53] P. H. Frampton, P. Q. Hung, and M. Sher, *Quarks and Leptons Beyond the Third Generation*, *Phys. Rept.* **330** (2000).

[54] M. Maltoni, V. A. Novikov, L. B. Okun, A. N. Rozanov, and M. I. Vysotsky, *Extra quark-lepton generations and precision measurements*, *Phys. Lett. B* **476** (2000).

[55] S. P. Martin, *A Supersymmetry Primer*, 9709356v6.

[56] S. Dimopoulos and H. Georgi, *Softly Broken Supersymmetry and SU(5)*, *Nucl. Phys. B* **193** (1981).

[57] W. Altmannshofer, A. J. Buras, S. Gori, P. Paradisi, and D. M. Straub, *Anatomy and Phenomenology of FCNC and CPV Effects in SUSY Theories*, arXiv:0909.1333.

[58] A. H. Chamseddine, R. L. Arnowitt, and P. Nath, *Locally Supersymmetric Grand Unification*, *Phys.Rev.Lett.* **49** (1982).

[59] G. L. Kane, C. Kolda, L. Roszkowski, and J. D. Wells, *Study of Constrained Minimal Supersymmetry*, *Phys.Rev.D* **49** (1994).

[60] J.Ellis *et. al.*, *WMAP-Compliant Benchmark Surfaces for MSSM Higgs Bosons*, arXiv:0709.0098.

[61] R. Babier *et. al.*, *R-Parity-violating supersymmetry*, *Phys. Rep.* **420** (2005).

[62] http://edu.pyhajoki.fi/lukiouusi/Oppiaineet/Fysiikka/cern2006/exercises/keyhole/sp/theory/main-9.html.

[63] J. E. Kim and H. P. Nilles, *The mu Problem and the Strong CP Problem*, *Phys. Lett. B* **138** (1984).

[64] M. Maniatis, *The Next-to-Minimal Supersymmetric extension of the Standard Model reviewed*, arXiv:0906.0777.

[65] A. Djouadi, U. Ellwanger, and A. Teixeira, *The constrained next-to-minimal supersymmetric standard model*, *Phys.Rev.Lett.* **101** (2008).

[66] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, *(De)Constructing Dimensions*, *Phys. Rev. Lett.* **86** (2001).

[67] H. C. Cheng and I. Low, *TeV Symmetry and the Little Hierarchy Problem*, *JHEP* **0309** (2003).

[68] H. C. Cheng and I. Low, *Little Hierarchy, Little Higgses, and a Little Symmetry*, *JHEP* **0408** (2004).

[69] S. Dynamics and E. S. Breaking, *Christopher T. Hill and Elizabeth H. Simmons*, *Phys.Rept.* **381** (2003).

[70] T. Hill, *"Topcolor"; Invited talk presented at "Strongly Coupled Gauge Theories"*, 9702320v1.

[71] T. Kaluza, *On the Unity Problem of Physics*, *Sitzungseber. Press. Akad. Wiss. Phys. Math. Klasse* **996** (1921).

[72] O. Klein, *Quantentheorie und fünfdimensionale relativitätstheorie*, *Z. Phys. A-Hadron Nucl.* **37** (1926).

[73] L. Randall and R. Sundrum, *A Large Mass Hierarchy from a Small Extra Dimension*, *Phys. Rev. Lett.* **83** (1999).

[74] T. Gherghetta and A. Pomarol, *Bulk Fields and Supersymmetry in a Slice of AdS*, *Nucl. Phys. B* **586** (2000).

[75] B. Meadows *et. al.*, *The impact of SuperB on avour physics*, arXiv:1109.5028.

[76] O. Buchmueller *et. al.*, *Higgs and Supersymmetry*, 1112.3564.

[77] www.cern.ch.

[78] *LHC to run at 4 TeV per beam in 2012*, . http://press.web.cern.ch/press/PressReleases/Releases2012/PR01.12E.html.

[79] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008).

[80] www.lhc-closer.es.

[81] **ATLAS** Collaboration, *Combined search for the Standard Model Higgs boson using up to 4.9 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC*, *Phys. Lett. B* **710** (2012).

[82] **CMS** Collaboration, *Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Lett. B* **710** (2012).

[83] https://indico.cern.ch/getFile.py/access?contribId=0&resId=0&materialId=slides&confId=197461.

[84] https://indico.cern.ch/getFile.py/access?contribId=1&resId=1&materialId=slides&confId=197461.

[85] **ATLAS** Collaboration, *ATLAS : Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN*, *CERN-LHCC* **94-43** (1994).

[86] **CMS** Collaboration, *Technical Proposal*, *CERN-LHCC* **94-38** (1994).

[87] **ALICE** Collaboration, *A Large Ion Collider Experiment - Technical Proposal*, *CERN-LHCC* **95-71** (1995).

[88] **LHCb** Collaboration, *LHCb Technical Proposal*, *CERN-LHCC* **98-4** (1998).

[89] **LHCb** Collaboration, *Roadmap for selected key measurements of LHCb*, *LHCb-PUB-2009-029* (2009).

[90] **LHCb** Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008).

[91] **LHCb** Collaboration, *LHCb magnet technical design report*, *CERN-LHCC* **2000-007** (2000).

[92] **LHCb** Collaboration, *LHCb VELO Technical Design Report*, *CERN-LHCC* **2001-011** (2001).

[93] **LHCb** Collaboration, *LHCb Reoptimized Detector Design and Performance Technical Design Report*, *CERN-LHCC* **2003-03** (2003).

[94] **LHCb** Collaboration, *LHCb Inner Tracker TDR*, *CERN-LHCC* **2002-029** (2002).

[95] **LHCb** Collaboration, *LHCb Outer Tracker TDR*, *CERN-LHCC* **2001-024** (2001).

[96] **LHCb** Collaboration, *LHCb RICH TDR*, *CERN-LHCC* **2000-050** (2000).

[97] **LHCb** Collaboration, *LHCb Calorimeters TDR*, *CERN-LHCC* **2000-036** (2000).

[98] **LHCb** Collaboration, *LHCb Muon System TDR*, *CERN-LHCC* **2001-010** (2001).

[99] P. Mato, *GAUDI-Architecture design document*, *LHCb note 98-064*.

[100] http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/moore/.

[101] `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/`.

[102] M. Deissenroth, *Experience with LHCb alignment software on first data*, *CERN-LHCb-CONF-2009-009* (2009).

[103] **LHCb** Collaboration, *Overview of LHCb alignment*, *1st LHC Detector Alignment Workshop* (2006).

[104] `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/alignment/`.

[105] `https://twiki.cern.ch/twiki/bin/view/LHCb/DaVinci`.

[106] `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/erasmus/`.

[107] I. Bird *et. al.*, *LHC Computing Grid TDR*, *CERN-LHCC* **2005-024** (2005).

[108] `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/`.

[109] http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/panoramix/.

[110] **LHCb** Collaboration, *Computing TDR*, *CERN-LHCC* **2005-019** (2005).

[111] **LHCb** Collaboration, *LHCb Trigger System TDR*, *CERN-LHCC* **2003-031** (2003).

[112] **LHCb** Collaboration, H. Ruiz-Perez, *The LHCb trigger: algorithms and performance*, *CERN-LHCb-CONF-2009-004* (2009).

[113] M. S. A. Martin Sanchez, P. Robbe, *Performances of the LHCb L0 Calorimeter Trigger*, *LHCb-PUB-2011-026* (2011).

[114] R. Aaij and J. Albrecht, *Muon triggers in the High Level Trigger of LHCb*, *LHCb-PUB-2011-017* (2011).

[115] V. V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers*, *LHCb-PUB-2011-016* (2011).

[116] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, *Nuclear Instruments and Methods in Physics Research A* **262** (1987).

[117] E. Bos, *Reconstruction of charged particles in the LHCb experiment*, *CERN-THESIS-2010-049* (2005).

[118] A. Jaeger, P. Seyfert, M. D. Cian, J. van Tilburg, and S. Hansmann-Menzemer, *Measurement of the track finding efficiency*, *LHCb-PUB-2011-025* (2012).

[119] M. Krasowski *et. al.*, *Primary vertex reconstruction*, *CERN-LHCb-2007-011* (2007).

[120] **LHCb Collaboration** Collaboration, *Measurement of the $B_s^0 - \bar{B}_s^0$ oscillation frequency Δm_s in $B_s^0 \to D_s^-(3)\pi$ decays*, *Phys.Lett.* **B709** (2012) 177–184, [arXiv:1112.4311].

[121] H. Terrier and I. Belyaev, *Electron particle identification with LHCb calorimeter system for 2011 data taking period*, *LHCb-2003-092* (2003).

[122] D. Golubkov and V. Egorychev, *Particle Identification with LHCb calorimeters*, *LHCb-INT-2011-052* (2011).

[123] I. Belyaev, V. Egorychev, and D. Golubkov, *Study of π^0/γ reconstruction efficiency with 2011 data*, *LHCb-INT-2012-001* (2012).

[124] G. Lanfranchi *et. al.*, *The Muon Identification procedure of the LHCb experiment for the first data*, *LHCb-PUB-2009-013*.

[125] X. Cid *et. al.*, *A Muon Identification procedure for LHCb with Kalman Filter*, *LHCb-INT-2010-052* (2010).

[126] X. Cid *et. al.*, *Performance of the LHCb Muon Identification*, *LHCb-INT-2012-016* (2012).

[127] R. Muresan, *LHCb - Minimum Bias physics plans and MC framework*, *7th Meeting of the MCnet European Network* (2010).

[128] T. S. S. Mrenna and P. Skands, *PYTHIA 6.4: Physics and manual*, *JHEP* **05** (2006).

[129] http://www.slac.stanford.edu/~lange/EvtGen/.

[130] J. Allison *et. al.*, *GEANT 4, toolkit for the simulation of the passage of particles through matter*, *Nuclear Science* **53** (2006).

[131] T. W. Anderson. Third ed., 2003.

[132] W. Altmannshofer *et. al.*, *Symmetries and Asymmetries of $B^0 \to K^{*0}\mu^+\mu^-$ Decays in the Standard Model and Beyond*, *JHEP* **0901** (2009).

[133] C. Bobeth, G. Hiller, and G. Piranishvili, *CP Asymmetries in bar $B \to \bar{K}^{*0}(\to \bar{K}\pi)l^+l^-$ and Untagged $\bar{B}_s, B_s \to \phi(\to K^+K^-)l^+l^-$ Decays at NLO*, *JHEP* **0807** (2008).

[134] **LHCb** Collaboration, *Differential branching fraction and angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay*, *LHCb-CONF-2012-008* (2012).

[135] A. Hocker *et. al.*, *TMVA: Toolkit for multivariate data analysis*, *PoS ACAT* **184** (2007).

[136] F. Mahmoudi, *Direct and indirect searches for New Physics*, arXiv:1205.3099.

[137] **LHCb** Collaboration, *Measurement of the isospin asymmetry in $B \to K^{(*)+-}$ decays*, arXiv:1205.3422.

[138] **LHCb** Collaboration, *First observation of $B^+ \to \pi^+\mu^+\mu^-$, LHCb-CONF-2012-006* (2012).

[139] **LHCb** Collaboration, *Measurement of the direct CP asymmetry in the $B^0 to K^{*0}\gamma$ decay, LHCb-CONF-2012-004* (2012).

[140] **LHCb** Collaboration, *Search for the lepton flavour violating decay $\tau^- \to \mu^+\mu^-\mu^-$, LHCb-CONF-2012-015* (2012).

[141] **LHCb** Collaboration, *Search for the rare decays $B_s^0 \to \mu^+\mu^-\mu^+\mu^-$ and $B^0 \to \mu^+\mu^-\mu^+\mu^-$, LHCb-CONF-2012-010* (2012).

[142] **LHCb** Collaboration, *Searches for Majorana neutrinos in B^- decays*, *Phys. Rev. D* **85** (2012).

[143] **LHCb** Collaboration, *Search for the $D^0 \to \mu^+\mu^-$ decay with 0.9 fb^1 at LHCb, LHCb-CONF-2012-005* (2012).

[144] **D0** Collaboration, *Search for the rare decay $B_s^0 \to \mu^+\mu^-$*, *Phys. Lett. B* **693** (2010).

[145] **CDF** Collaboration, *Search for $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ Decays with CDF II*, *Phys. Rev. Lett.* **107** (2011).

[146] **CMS** Collaboration, *Search for $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ decays in pp Collisions at 7 TeV*, *Phys. Rev. Lett.* **107** (2011).

[147] **LHCb** Collaboration, *Search for the rare decays $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$, Phys. Lett. B* **708** (2012).

[148] **LHCb** Collaboration, *Search for the Rare Decays $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$, Phys.Lett. B* **699** (2011).

[149] H. Miyake, *Heavy Flavor Physics at the Tevatron*, *XXVI Rencontres de Physique de la Vallée d'Aoste* (2012).
http://www-cdf.fnal.gov/physics/talks_transp/2012/5-Miyake.pdf.

[150] **LHCb** Collaboration, *Measurement of $\sigma(pp \to b\bar{b}\ X)$ at $\sqrt(s)$=7 TeV in the forward region*, *Phys. Lett. B* **694** (2010).

[151] T. Skwarnicki, *A Study of the Radiative Cascade Transitions Between the Υ and Υ' resonances*, *DESY F31-86-02* (1986).

[152] A. L. Read, *Presentation of search results: The cl_s technique*, *J. Phys.* **G28** (2002) 2693–2704.

[153] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth.* **A434** (1999) 435–443, [hep-ex/9902006].

[154] https://twiki.cern.ch/twiki/bin/view/LHCb/CondDBHowTo.

[155] https://lhcb-comp.web.cern.ch/lhcb-comp/Frameworks/DetDesc/default.htm.

[156] http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/lhcb/.

[157] http://heptools.inp.demokritos.gr/node/1021.

[158] `http://projects.hepforge.org/lhapdf/`.

[159] `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/`.

[160] `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/STATISTICS/MC10STAT/RD-WG/Generation_Beam3500GeV-md100-MC10-nu2,5.html#10012005`.

[161] **LHCb** Collaboration, *Measurement of J/ψ production in pp collisions at $\sqrt{s}$ = 7 TeV*, *Eur. Phys. J.* **C71** (2011).

[162] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, *Ann. Math. Statist.* **22** (1951).

[163] S. Kullback, *Letter to Editor: the Kullback-Leibler Distance*, *The American Statistician* **41** (1987).

[164] M. Needham, *Clone Track identification using the Kullback-Liebler distance*, *LHCb-2008-002* (2008).

[165] C. Adrover *et. al.*, *Search for the rare decays $B^0_{(s)} \to \mu^+\mu^-$ with 370 pb^{-1} at LHCb*, *LHCb-ANA-2011-078* (2011).

[166] K. Akiba *et. al.*, *Results on Muon identification efficiency with 2011 data at LHCb*, *LHCb-INT-2011-045* (2011).

[167] **CDF** Collaboration, *Observation of the B_c meson in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys. Rev. Lett.* **81** (1998).

[168] C. Adrover *et. al.*, *Search for the rare decays $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ with the LHCb experiment*, *LHCb-ANA-2011-078* (2011).

[169] M. P. Terrin and G. Mancinelli, *Optimization of the binning of the discriminating variables used in the computation of the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ upper limits with the modified frequentist approach*, *LHCb-INT-2012-003* (2012).

[170] R. Barlow, *Asymmetric Errors*, `physics/0401042v1`.

[171] A. Carbone *et. al.*, *Invariant mass line shape of $B \to h^+h^-$ decays at LHCb*, *LHCb-PUB-2009-031* (2009).

[172] C. Adrover *et. al.*, *Search for the rare decays $B^0_{(s)} \to \mu^+\mu^-$ with the LHCb experiment*, *LHCb-ANA-2011-007* (2011).

[173] C. Adrover *et. al.*, *Search for the rare decays $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ with 1.02 fb^{-1} at LHCb*, *LHCb-ANA-2011-102*.

[174] L. Carson *et. al.*, *Measurement of the relative yields of the decay modes $B_{d,s} \to D_{d,s}\pi$ and $B_{d,s} \to D_{d,s}K$ and determination of f_d/f_s for 7* TeV *pp collisions*, *LHCb-ANA-2010-010* (2010).

[175] M. Artuso *et. al.*, *Measurement of b-hadron production fractions in 7*TeV *centre-of-mass energy pp collisions*, *LHCb-ANA-2010-015* (2010).

[176] **LHCb** Collaboration, *Average f_s/f_d b–hadron production fraction for 7 TeV pp collisions*, *CERN-LHCb-CONF-2011-034* (2011).

[177] `https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbTrackingEfficiencies`.

[178] K. Akiba *et. al.*, *Measurement of Δm_d in the decay $B_d \to D(K\pi\pi)\pi$ and calibration of the combined opposite side taggers*, *LHCb-ANA-2011-004* (2011).

[179] E. Lopez Asamar *et. al.*, *Measurement of trigger efficiencies and biases*, *LHCb-PUB-2007-073* (2007).

[180] D. M. Straub, *New physics correlations in rare decays*, `arXiv:1012.3893`.

[181] K. Agashe and C. D. Carone, *Supersymmetric Flavor Models and the $B \to \phi K^0_S$ Anomaly*, *Phys. Rev. D* **68** (2003).

[182] G. G. Ross, L. Velasco-Sevilla, and O. Vives, *Spontaneous CP violation and Non-Abelian Family Symmetry in SUSY*, *Nucl. Phys. B* **692** (2004).

[183] S. Antusch, S. F. King, and M. Malinsky, *Solving the SUSY Flavour and CP Problems with SU(3) Family Symmetry*, *JHEP* **0806** (2008).

[184] L. J. Hall and H. Murayama, *A Geometry of the Generations*, *Phys. Rev. Lett.* **75** (1995).

[185] O. Buchmueller *et. al.*, *The CMSSM and NUHM1 in Light of 7 TeV LHC, $B^0_s \to \mu^+\mu^-$ and XENON100 Data*, `arXiv:1207.7315`.

[186] **XENON100** Collaboration, E. Aprile *et. al.*, *Dark Matter Results from 225 Live Days of XENON100 Data*, `arXiv:1207.5988`.

[187] P. Bechtle *et. al.*, *Constrained Supersymmetry after two years of LHC data: a global view with Fittino*, `arXiv:1204.4199`.

[188] A. Fowlie *et. al.*, *The CMSSM Favoring New Territories: The Impact of New LHC Limits and a 125 GeV Higgs*, `arXiv:1206.0264`.

[189] A. Akeroyd, F. Mahmoudi, and D. Martinez, *The decay $B^0_s \to \mu^+\mu^-$: updated SUSY constraints and prospects*, *JHEP* **1112** (2011).

[190] **CMS** Collaboration, S. Chatrchyan *et. al.*, *Search for Supersymmetry at the LHC in Events with Jets and Missing Transverse Energy*, *Phys. Rev. Lett.* **107** (2011).

[191] **CMS** Collaboration, *Search for supersymmetry with the razor variables at CMS*, *CMS-PAS-SUS-12-005* (2012).

[192] **ATLAS** Collaboration, *Search for the decay $B^0_s \to \mu^+\mu^-$ with the ATLAS detector*, `arXiv:1204.0735`.

[193] **CMS** Collaboration, *Search for $B^0_s \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays*, *JHEP* **1204** (2012).

[194] **LHCb** Collaboration, *Search for the rare decays $B^0_{(s)} \to \mu^+\mu^-$ at the LHC with the ATLAS, CMS and LHCb experiments*, *LHCb-CONF-2012-017* (2012).

[195] S. Gjesdal, G. Presser, P. Steffen, J. Steinberger, F. Vannucci, *et. al.*, *Search for the decay $K_s \to 2\mu$*, *Phys.Lett. B* **44** (1973) 217–220.

[196] D. Martinez, *Use of 2009 real data V0s to check some $B^0_s \to \mu^+\mu^-$ key points*, *CERN-LHCb-INT-2010-010* (2010).

[197] `https://twiki.cern.ch/twiki/bin/view/LHCb/CommonParticles`.

[198] J. Podolanski and R. Armenteros, *Analysis of V-events*, *Phil. Mag.* **45** (1954) 13.

[199] C. Jones, *Charged Particle ID*, *LHCb: 47th Analysis and Software week.* `https://indico.cern.ch/getFile.py/access?contribId=4&sessionId=0&resId=0&materialId=slides&confId=159519`.

[200] M. Gandelman and E. Polycarpo, *The Performance of the LHCb Muon Identification Procedure*, *LHCb-2007-145*.

[201] N. Gauvin *et. al.*, *Search for (Higgs-like) bosons decaying into long-lived exotic particles*, *LHCb-ANA-2012-038*.

[202] M. O. Bettler, X. Cid, and D. Martinez, *Search for $K^0_S \to \mu^+\mu^-$*, *LHCb-ANA-2011-101* (2012).

[203] V. Cirigliano, G. Ecker, H. Neufeld, A. Pich, and J. Portoles, *Kaon Decays in the Standard Model*, *Rev.Mod.Phys.* **84** (2012) 399, [arXiv:1107.6001].

[204] **LHCb Collaboration** Collaboration, R. Aaij *et. al.*, *Measurement of V^0 production ratios in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV*, *JHEP* **1108** (2011) 034, [arXiv:1107.0882].

[205] J. Albrecht *et. al.*, *Updated measurement of the J/Ψ production cross-section in LHCb*, *LHCb-ANA-2010-012*.